GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:33:46 ; Search time 70.07 Seconds
(without alignments)
711.845 Million cell updates/sec

Title: US-09-500-376-5
Perfect score: 1982
Sequence: 1 AVTTSVIDNILSKIENEYEV.....SNFLGISFLLILMLILYSFI 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1824.5	92.1	400	5	Q03999	Q03999 plasmodium
2	1824.5	92.1	539	5	Q25966	Q25966 plasmodium
3	1824.5	92.1	539	5	Q25976	Q25976 plasmodium
4	1824.5	92.1	539	5	Q25984	Q25984 plasmodium
5	1824.5	92.1	539	5	Q9TYG1	Q9tyg1 plasmodium
6	1821.5	91.9	539	5	Q25972	Q25972 plasmodium
7	1821.5	91.9	539	5	Q25981	Q25981 plasmodium
8	1818.5	91.8	539	5	Q25971	Q25971 plasmodium
9	1818.5	91.8	539	5	Q25973	Q25973 plasmodium
10	1167.5	58.9	651	5	Q25924	Q25924 plasmodium
11	1164.5	58.8	569	5	Q25969	Q25969 plasmodium
12	1164.5	58.8	569	5	Q25974	Q25974 plasmodium
13	1164.5	58.8	569	5	Q25975	Q25975 plasmodium
14	1164.5	58.8	569	5	Q25977	Q25977 plasmodium
15	1164.5	58.8	569	5	Q25979	Q25979 plasmodium
16	1164.5	58.8	1694	5	Q9TZT5	Q9tzt5 plasmodium
17	1164.5	58.8	1694	5	Q9NHX1	Q9nhx1 plasmodium
18	1164.5	58.8	1704	5	Q9TZT4	Q9tzt4 plasmodium
19	1163.5	58.7	569	5	Q25983	Q25983 plasmodium
20	1151.5	58.1	569	5	Q25967	Q25967 plasmodium
21	1145.5	57.8	569	5	Q25970	Q25970 plasmodium
22	1145.5	57.8	569	5	Q25980	Q25980 plasmodium
23	1145.5	57.8	569	5	Q25982	Q25982 plasmodium
24	1145	57.8	570	5	Q25968	Q25968 plasmodium
25	1145	57.8	570	5	Q9TYG2	Q9tyg2 plasmodium
26	1144.5	57.7	569	5	Q25978	Q25978 plasmodium
27	1134	57.2	1720	5	Q25922	Q25922 plasmodium
28	1130	57.0	652	5	Q25923	Q25923 plasmodium
29	1069	53.9	373	5	Q25724	Q25724 plasmodium
30	1067.5	53.9	372	5	Q25725	Q25725 plasmodium
31	1066	53.8	373	5	Q25721	Q25721 plasmodium
32	1063.5	53.7	372	5	O43997	O43997 plasmodium
33	1063	53.6	373	5	O43995	O43995 plasmodium
34	1062.5	53.6	372	5	Q25726	Q25726 plasmodium
35	1059	53.4	373	5	Q25722	Q25722 plasmodium
36	1058	53.4	373	5	Q25723	Q25723 plasmodium
37	1048.5	52.9	372	5	Q25719	Q25719 plasmodium
38	1048.5	52.9	372	5	Q25720	Q25720 plasmodium
39	1048	52.9	373	5	Q25727	Q25727 plasmodium
40	1047.5	52.9	372	5	Q25717	Q25717 plasmodium
41	1045	52.7	373	5	O43996	O43996 plasmodium
42	1043	52.6	373	5	Q25728	Q25728 plasmodium
43	1042.5	52.6	372	5	Q25718	Q25718 plasmodium
44	1024.5	51.7	219	5	Q9UB87	Q9ub87 plasmodium
45	747	37.7	599	5	Q9NCN2	Q9ncn2 plasmodium

ALIGNMENTS

RESULT 1
Q03999
ID Q03999 PRELIMINARY; PRT; 400 AA.
AC Q03999;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE MEROZOITE SURFACE ANTIGEN PRECURSOR 1 (PMMSA) (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Blackman M.J., Ling I.T., Nicholls S.C., Holder A.A.;
RL Submitted (XXX-1991) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: COULD BE ATTACHED TO THE MEMBRANE BY A GPI-
CC ANCHOR.
CC -!- MISCELLANEOUS: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF
CC 83K, 42K & 19K ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
DR EMBL; M64681; AAA29709.1; -.
DR InterPro; IPR000561; -.
DR InterPro; IPR001245; -.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT NON_TER 1 1
FT SIGNAL <1 24 POTENTIAL.
FT CHAIN 25 400 83K MEROZOITE SURFACE ANTIGEN.
FT CHAIN 25 286 42K MEROZOITE SURFACE ANTIGEN.
FT CHAIN 287 400 19K MEROZOITE SURFACE ANTIGEN.
FT TRANSMEM 383 400 MEMBRANE ANCHOR.
SQ SEQUENCE 400 AA; 45824 MW; 537F075058626AC2 CRC64;

Query Match 92.1%; Score 1824.5; DB 5; Length 400;
Best Local Similarity 93.6%; Pred. No. 1.5e-87;
Matches 353; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

Qy 1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60

```
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 84

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 144

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 204

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     205 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 264

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
          ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||||
Db     265 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 323

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     324 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 383

Qy     361 LGISFLLILMLILYSFI 377
          |||||||||||||||||
Db     384 LGISFLLILMLILYSFI 400


RESULT   2
Q25966
ID   Q25966         PRELIMINARY;   PRT;   539 AA.
AC   Q25966;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13357; BAA02618.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1       1
SQ   SEQUENCE   539 AA;  61144 MW;  7B7DE90C1D0ACDE7 CRC64;


  Query Match             92.1%;  Score 1824.5;  DB 5;  Length 539;
  Best Local Similarity   93.6%;  Pred. No. 2e-87;
  Matches  353;  Conservative    6;  Mismatches   17;  Indels    1;  Gaps    1;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
          ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||||
Db     404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 462

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     463 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy     361 LGISFLLILMLILYSFI 377
          |||||||||||||||||
Db     523 LGISFLLILMLILYSFI 539


RESULT   3
Q25976
ID   Q25976         PRELIMINARY;   PRT;   539 AA.
AC   Q25976;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13362; BAA02623.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1       1
SQ   SEQUENCE   539 AA;  61045 MW;  8BF9C64322E9A778 CRC64;


  Query Match             92.1%;  Score 1824.5;  DB 5;  Length 539;
  Best Local Similarity   93.6%;  Pred. No. 2e-87;
  Matches  353;  Conservative    6;  Mismatches   17;  Indels    1;  Gaps    1;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
          ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||||
Db     404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 462

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    463 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy    361 LGISFLLILMLILYSFI 377
          |||||||||||||||||
Db    523 LGISFLLILMLILYSFI 539


RESULT   4
Q25984
ID   Q25984         PRELIMINARY;   PRT;   539 AA.
AC   Q25984;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13356; BAA02617.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   539 AA;  61114 MW;  3788015F3127CB9E CRC64;


  Query Match             92.1%;  Score 1824.5;  DB 5;  Length 539;
  Best Local Similarity   93.6%;  Pred. No. 2e-87;
  Matches  353;  Conservative    6;  Mismatches   17;  Indels    1;  Gaps    1;

Qy      1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy    181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy    241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
          ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||||
Db    404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 462

Qy    301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    463 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy    361 LGISFLLILMLILYSFI 377
          |||||||||||||||||
Db    523 LGISFLLILMLILYSFI 539


RESULT   5
Q9TYG1
ID   Q9TYG1      PRELIMINARY;      PRT;   539 AA.
AC   Q9TYG1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13358; BAA02619.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   539 AA;  61075 MW;  C70C2E100EC4A101 CRC64;


  Query Match             92.1%;  Score 1824.5;  DB 5;  Length 539;
  Best Local Similarity   93.6%;  Pred. No. 2e-87;
  Matches  353;  Conservative    6;  Mismatches   17;  Indels    1;  Gaps    1;

Qy      1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy    181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy    241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
          ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||||
Db    404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 462

Qy    301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    463 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy    361 LGISFLLILMLILYSFI 377
          |||||||||||||||||
Db    523 LGISFLLILMLILYSFI 539


RESULT   6
Q25972
ID   Q25972         PRELIMINARY;   PRT;   539 AA.
AC   Q25972;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13360; BAA02621.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1       1
SQ   SEQUENCE   539 AA;  61076 MW;  72186F1412C57D1C CRC64;


  Query Match             91.9%;  Score 1821.5;  DB 5;  Length 539;
  Best Local Similarity   93.4%;  Pred. No. 2.9e-87;
  Matches  352;  Conservative    7;  Mismatches   17;  Indels    1;  Gaps    1;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
           ||||| |:|||||||  | : : : :     |    ::|||||||||||||||||||||||
Db     404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQE 462

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
           ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     463 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy     361 LGISFLLILMLILYSFI 377
           |||||||||||||||||
Db     523 LGISFLLILMLILYSFI 539


RESULT   7
Q25981
ID   Q25981         PRELIMINARY;   PRT;   539 AA.
AC   Q25981;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium..
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13363; BAA02624.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1       1
SQ   SEQUENCE   539 AA;  61046 MW;  3EED87473EE87B65 CRC64;


  Query Match             91.9%;  Score 1821.5;  DB 5;  Length 539;
  Best Local Similarity   93.4%;  Pred. No. 2.9e-87;
  Matches  352;  Conservative    7;  Mismatches   17;  Indels    1;  Gaps    1;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
           ||||| |:|||||||  | : : : :     |    ::|||||||||||||||||||||||
Db     404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQE 462

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
           ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     463 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy     361 LGISFLLILMLILYSFI 377
           |||||||||||||||||
Db     523 LGISFLLILMLILYSFI 539


RESULT   8
Q25971
ID   Q25971         PRELIMINARY;   PRT;   539 AA.
AC   Q25971;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13359; BAA02620.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1       1
SQ   SEQUENCE   539 AA;  61017 MW;  5CA9C651BB62B5E1 CRC64;


  Query Match             91.8%;  Score 1818.5;  DB 5;  Length 539;
  Best Local Similarity   93.4%;  Pred. No. 4.1e-87;
```

```
  Matches  352;  Conservative    6;  Mismatches   18;  Indels    1;  Gaps    1;

Qy      1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy    181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy    241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
          ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||||
Db    404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 462

Qy    301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
          | ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    463 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy    361 LGISFLLILMLILYSFI 377
          |||||||||||||||||
Db    523 LGISFLLILMLILYSFI 539


RESULT  9
Q25973
ID   Q25973         PRELIMINARY;   PRT;   539 AA.
AC   Q25973;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13361; BAA02622.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   539 AA;  61047 MW;  105C2E02974FB398 CRC64;


  Query Match             91.8%;  Score 1818.5;  DB 5;  Length 539;
  Best Local Similarity   93.4%;  Pred. No. 4.1e-87;
  Matches  352;  Conservative    6;  Mismatches   18;  Indels    1;  Gaps    1;

Qy      1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy    181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy    241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
          ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||||
Db    404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 462

Qy    301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
          | ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    463 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy    361 LGISFLLILMLILYSFI 377
          |||||||||||||||||
Db    523 LGISFLLILMLILYSFI 539


RESULT  10
Q25924
ID   Q25924      PRELIMINARY;      PRT;   651 AA.
AC   Q25924;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MEROZOITE SURFACE ANTIGEN 1 (FRAGMENT).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RO-71;
RX   MEDLINE=92275047; PubMed=1592091;
RA   Olafsson P., Matile H., Certa U.;
RT   "Plasmodium falciparum: the repetitive MSA-1 surface protein of the
RT   RO-71 isolate is recognized by mouse antibody against the
RT   nonrepetitive repeat block of RO-33.";
RL   Exp. Parasitol. 74:381-389(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RO-71;
RX   MEDLINE=95354793; PubMed=7628566;
RA   Tolle R., Bujard H., Cooper J.A.;
RT   "Plasmodium falciparum: variations within the C-terminal region of
RT   merozoite surface antigen-1.";
RL   Exp. Parasitol. 81:47-54(1995).
DR   EMBL; Z35329; CAA84558.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite.
FT   NON_TER       1      1
SQ   SEQUENCE   651 AA;  74134 MW;  AA2137E699255150 CRC64;


  Query Match             58.9%;  Score 1167.5;  DB 5;  Length 651;
  Best Local Similarity   57.9%;  Pred. No. 2.6e-53;
  Matches  226;  Conservative   59;  Mismatches   84;  Indels   21;  Gaps    3;

Qy      7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
          :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db    264 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 323

Qy     67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
          ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db    324 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 383

Qy    127 SEKYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVND 167
            ||| ||:|||| |                    |:| :  |:|||| ||||||  : :
Db    384 LAKYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVN 443
```

```
Qy    168 KIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYN 227
          ||| ::|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db    444 KIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDDT 503

Qy    228 HNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDE 287
            ::| | ||||:| :|   :::| |:: |  : : : :     |    |:|||||||||
Db    504 KKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPQNSGCFRHLDE 561

Qy    288 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCY 347
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    562 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 621

Qy    348 PLFDGIFCSSSNFLGISFLLILMLILYSFI 377
          ||||||||||||||||||||||||||||||
Db    622 PLFDGIFCSSSNFLGISFLLILMLILYSFI 651


RESULT  11
Q25969
ID   Q25969         PRELIMINARY;   PRT;   569 AA.
AC   Q25969;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13346; BAA02607.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64485 MW;  A26C45914AF01396 CRC64;


  Query Match             58.8%;  Score 1164.5;  DB 5;  Length 569;
  Best Local Similarity   57.7%;  Pred. No. 3.2e-53;
  Matches  225;  Conservative   60;  Mismatches   84;  Indels   21;  Gaps    3;

Qy      7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
          :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db    182 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 241

Qy     67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
          ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db    242 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 301

Qy    127 SEKYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVND 167
            ||| ||:|||| |                  |:| :  |:|||| ||||||  : :
Db    302 LAKYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVN 361

Qy    168 KIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYN 227
          ||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db    362 KIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDT 421

Qy    228 HNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDE 287
            ::| | ||||:| :|   :::| |:: |  : : : :     |    ::|||||||||
Db    422 KKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDE 479

Qy    288 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCY 347
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    480 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 539

Qy    348 PLFDGIFCSSSNFLGISFLLILMLILYSFI 377
          ||||||||||||||||||||||||||||||
Db    540 PLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT  12
Q25974
ID   Q25974         PRELIMINARY;   PRT;   569 AA.
AC   Q25974;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13348; BAA02609.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64457 MW;  A2691FE569E80396 CRC64;


  Query Match             58.8%;  Score 1164.5;  DB 5;  Length 569;
  Best Local Similarity   57.7%;  Pred. No. 3.2e-53;
  Matches  225;  Conservative   60;  Mismatches   84;  Indels   21;  Gaps    3;

Qy      7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
          :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db    182 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 241

Qy     67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
          ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db    242 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 301

Qy    127 SEKYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVND 167
            ||| ||:|||| |                  |:| :  |:|||| ||||||  : :
Db    302 LAKYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVN 361

Qy    168 KIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYN 227
          ||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db    362 KIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDT 421

Qy    228 HNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDE 287
            ::| | ||||:| :|   :::| |:: |  : : : :     |    ::|||||||||
Db    422 KKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDE 479

Qy    288 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCY 347
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    480 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 539

Qy    348 PLFDGIFCSSSNFLGISFLLILMLILYSFI 377
          ||||||||||||||||||||||||||||||
Db    540 PLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT  13
Q25975
ID   Q25975         PRELIMINARY;   PRT;   569 AA.
AC   Q25975;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
```

```
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13349; BAA02610.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1       1
SQ   SEQUENCE   569 AA;  64459 MW;  ED4284B2867C9703 CRC64;


  Query Match             58.8%;  Score 1164.5;  DB 5;  Length 569;
  Best Local Similarity   57.7%;  Pred. No. 3.2e-53;
  Matches  225;  Conservative   60;  Mismatches   84;  Indels   21;  Gaps    3;

Qy       7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
           :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db     182 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 241

Qy      67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
           ||||: :| ::|: |:::| :| || |::  | || |||:|:: || || : : ||: :
Db     242 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 301

Qy     127 SEKYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVND 167
             ||| ||:|||| |                    |:| :  |:|||| ||||||  : :
Db     302 LAKYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVN 361

Qy     168 KIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYN 227
           ||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db     362 KIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDT 421

Qy     228 HNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDE 287
             ::| | ||||:| :|   :::| |:: |  : : : :     |    ::|||||||||
Db     422 KKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDE 479

Qy     288 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCY 347
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     480 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 539

Qy     348 PLFDGIFCSSSNFLGISFLLILMLILYSFI 377
           |||||||||||||||||||||||||||||
Db     540 PLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT  14
Q25977
ID   Q25977         PRELIMINARY;   PRT;   569 AA.
AC   Q25977;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13350; BAA02611.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1       1
SQ   SEQUENCE   569 AA;  64485 MW;  A26C45914AF01396 CRC64;


  Query Match             58.8%;  Score 1164.5;  DB 5;  Length 569;
  Best Local Similarity   57.7%;  Pred. No. 3.2e-53;
  Matches  225;  Conservative   60;  Mismatches   84;  Indels   21;  Gaps    3;

Qy       7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
           :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db     182 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 241

Qy      67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
           ||||: :| ::|: |:::| :| || |::  | || |||:|:: || || : : ||: :
Db     242 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 301

Qy     127 SEKYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVND 167
             ||| ||:|||| |                    |:| :  |:|||| ||||||  : :
Db     302 LAKYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVN 361

Qy     168 KIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYN 227
           ||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db     362 KIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDT 421

Qy     228 HNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDE 287
             ::| | ||||:| :|   :::| |:: |  : : : :     |    ::|||||||||
Db     422 KKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDE 479

Qy     288 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCY 347
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     480 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 539

Qy     348 PLFDGIFCSSSNFLGISFLLILMLILYSFI 377
           |||||||||||||||||||||||||||||
Db     540 PLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT  15
Q25979
ID   Q25979         PRELIMINARY;   PRT;   569 AA.
AC   Q25979;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13352; BAA02613.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1       1
SQ   SEQUENCE   569 AA;  64457 MW;  A2691FE569E80396 CRC64;


  Query Match             58.8%;  Score 1164.5;  DB 5;  Length 569;
  Best Local Similarity   57.7%;  Pred. No. 3.2e-53;
  Matches  225;  Conservative   60;  Mismatches   84;  Indels   21;  Gaps    3;
```

```
Qy       7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
           :|||||  ||||:|:||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db     182 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 241

Qy      67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
           ||||: :| ::|: |:::| :| || |::  | || |||:|:: || || : : ||: :
Db     242 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 301

Qy     127 SEKYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVND 167
             ||| ||:|||| |                  |:| :  |:|||| |||||  : :
Db     302 LAKYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVN 361

Qy     168 KIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYN 227
           ||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db     362 KIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDT 421

Qy     228 HNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDE 287
             ::| | ||||:| :|   :::| |:: |  : : : :     |    ::|||||||||
Db     422 KKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDE 479

Qy     288 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCY 347
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     480 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 539

Qy     348 PLFDGIFCSSSNFLGISFLLILMLILYSFI 377
           |||||||||||||||||||||||||||||
Db     540 PLFDGIFCSSSNFLGISFLLILMLILYSFI 569
```

Search completed: August 8, 2001, 12:33:47
Job time: 275 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:32:38 ; Search time 24.79 Seconds
(without alignments)
520.949 Million cell updates/sec

Title: US-09-500-376-5
Perfect score: 1982
Sequence: 1 AVTTSVIDNILSKIENEYEV.....SNFLGISFLLILMLILYSFI 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1824.5	92.1	1630	1	MSP1_PLAFK	P04932 plasmodium
2	1824.5	92.1	1639	1	MSP1_PLAFW	P04933 plasmodium
3	1164	58.7	1726	1	MSP1_PLAFC	P04934 plasmodium
4	1163	58.7	1726	1	MSP1_PLAFP	P50495 plasmodium
5	1156.5	58.4	1682	1	MSP1_PLAF3	P19598 plasmodium
6	1145	57.8	1701	1	MSP1_PLAFF	P13819 plasmodium
7	1145	57.8	1701	1	MSP1_PLAFM	P08569 plasmodium
8	588.5	29.7	1772	1	MSP1_PLAYO	P13828 plasmodium
9	142	7.2	1251	1	RBP2_PLAVB	Q00799 plasmodium
10	139.5	7.0	377	1	Y704_METJA	Q58115 methanococc
11	136.5	6.9	2198	1	YLJ2_CAEEL	P34367 caenorhabdi
12	135	6.8	1102	1	RPOP_AGABT	P33539 agaricus bi
13	134	6.8	1150	1	IRR1_YEAST	P40541 saccharomyc
14	133	6.7	748	1	Y875_METJA	Q58285 methanococc
15	133	6.7	864	1	CHEA_BORBU	Q44737 borrelia bu
16	133	6.7	1030	1	Y018_MYCPN	P75093 mycoplasma
17	130.5	6.6	881	1	LHS1_YEAST	P36016 saccharomyc
18	130	6.6	1162	1	BXEN_CLOBU	Q06366 clostridium
19	130	6.6	1196	1	BXCN_CLOBO	P46081 clostridium
20	129.5	6.5	1956	1	ATX1_PLAFA	Q04956 plasmodium
21	129	6.5	756	1	Y328_MYCGE	Q49419 mycoplasma
22	129	6.5	1162	1	BXEN_CLOBO	P46082 clostridium
23	129	6.5	1169	1	EX5B_BORBU	O51578 borrelia bu
24	128.5	6.5	867	1	DPO1_RICPR	O05949 rickettsia
25	127	6.4	766	1	YS83_CAEEL	Q09622 caenorhabdi
26	126.5	6.4	944	1	NUF1_YEAST	P32380 saccharomyc
27	126	6.4	703	1	HS83_PHANI	P51819 pharbitis n
28	125.5	6.3	655	1	YKDA_MYCCA	P45615 mycoplasma
29	125	6.3	1038	1	CIN8_YEAST	P27895 saccharomyc
30	125	6.3	1901	1	YCF1_TOBAC	P12222 nicotiana t
31	124.5	6.3	2869	1	RBP1_PLAVB	Q00798 plasmodium
32	123.5	6.2	989	1	SERA_PLAFG	P13823 plasmodium
33	122.5	6.2	749	1	MAD1_YEAST	P40957 saccharomyc
34	122	6.2	3135	1	S230_PLAFO	Q08372 plasmodium
35	121	6.1	632	1	Y242_MYCPN	P75440 mycoplasma
36	120.5	6.1	629	1	PR39_YEAST	P39682 saccharomyc
37	119.5	6.0	971	1	Y228_BORBU	O51246 borrelia bu
38	119.5	6.0	1435	1	EBA1_PLAFC	P19214 plasmodium
39	119.5	6.0	1790	1	USO1_YEAST	P25386 saccharomyc
40	119.5	6.0	2014	1	YJU7_YEAST	P39526 saccharomyc
41	119	6.0	404	1	K1MB_HUMAN	Q14525 homo sapien
42	118.5	6.0	515	1	YD56_SCHPO	Q10310 schizosacch
43	118.5	6.0	839	1	YNE1_YEAST	P53959 saccharomyc
44	118.5	6.0	845	1	SCP1_MESAU	Q60563 mesocricetu
45	118	6.0	453	1	YKP9_KLULA	P05475 kluyveromyc

ALIGNMENTS

RESULT 1
MSP1_PLAFK
ID MSP1_PLAFK STANDARD; PRT; 1630 AA.
AC P04932;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
RT Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RA Pan W., Tolle R., Bujard H.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC ----------

CC ----------
DR EMBL; X03371; CAA27070.1; -.
DR PIR; A25120; SAZQK1.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.

FT	SIGNAL	1 - 19	POTENTIAL.
FT	CHAIN	20 - 1630	MEROZOITE SURFACE PROTEIN 1.
FT	DOMAIN	67 - 84	TRIPEPTIDE SG(TP) REPEAT.
FT	TRANSMEM	1614 - 1630	MEMBRANE ANCHOR.
FT	CARBOHYD	97	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	259	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	755	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	759	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	774	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	835	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	911	N-LINKED (GLCNAC. . .) (POTENTIAL).

```
FT   CARBOHYD    955    955       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1049   1049       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1156   1156       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1165   1165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1436   1436       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1517   1517       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1630 AA;  187289 MW;  ADBDEC3CE0A46322 CRC64;


  Query Match           92.1%;  Score 1824.5;  DB 1;  Length 1630;
  Best Local Similarity  93.6%;  Pred. No. 2.5e-90;
  Matches  353;  Conservative    6;  Mismatches   17;  Indels    1;  Gaps    1;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1255 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 1314

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1315 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 1374

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1375 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 1434

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1435 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 1494

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
           ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||
Db    1495 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 1553

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
           ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1554 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 1613

Qy     361 LGISFLLILMLILYSFI 377
           |||||||||||||||||
Db    1614 LGISFLLILMLILYSFI 1630


RESULT   2
MSP1_PLAFW
ID   MSP1_PLAFW     STANDARD;      PRT;  1639 AA.
AC   P04933;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P195).
GN   MSP-1.
OS   Plasmodium falciparum (isolate Wellcome).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5848;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86014355; PubMed=2995820;
RA   Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA   Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA   Freeman R.R.;
RT   "Primary structure of the precursor to the three major surface
RT   antigens of Plasmodium falciparum merozoites.";
RL   Nature 317:270-273(1985).
RN   [2]
RP   REVISIONS.
RA   Holder A.A.;
RL   Submitted (MAR-1991) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X02919; CAA26676.1; -.
DR   PIR; A24594; A24594.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1639       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD   116    116       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   268    268       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   764    764       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   768    768       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   783    783       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   844    844       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   920    920       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   964    964       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1058   1058       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1165   1165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1174   1174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1445   1445       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1526   1526       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1639 AA;  187618 MW;  2C255B6616C87F6E CRC64;


  Query Match           92.1%;  Score 1824.5;  DB 1;  Length 1639;
  Best Local Similarity  93.6%;  Pred. No. 2.6e-90;
  Matches  353;  Conservative    6;  Mismatches   17;  Indels    1;  Gaps    1;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1264 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 1323

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1324 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 1383

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1384 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 1443

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1444 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 1503

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
           ||||| |:|||||||  | : : : :     |    ]:||||||||||||||||||||
Db    1504 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 1562

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
           ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1563 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 1622

Qy     361 LGISFLLILMLILYSFI 377
           |||||||||||||||||
Db    1623 LGISFLLILMLILYSFI 1639


RESULT   3
MSP1_PLAFC
ID   MSP1_PLAFC     STANDARD;      PRT;  1726 AA.
AC   P04934;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
```

```
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P195).
GN   MSP-1.
OS   Plasmodium falciparum (isolate Camp / Malaysia).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5835;
RN   [1]
RP   SEQUENCE OF 1-1103 FROM N.A.
RX   MEDLINE=86205236; PubMed=3517809;
RA   Weber J.L., Leininger W.M., Lyon J.A.;
RT   "Variation in the gene encoding a major merozoite surface antigen of
RT   the human malaria parasite Plasmodium falciparum.";
RL   Nucleic Acids Res. 14:3311-3323(1986).
RN   [2]
RP   SEQUENCE OF 1104-1726 FROM N.A.
RX   MEDLINE=88143999; PubMed=3278296;
RA   Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
RT   "Merozoite surface protein sequence from the Camp strain of the human
RT   malaria parasite Plasmodium falciparum.";
RL   Nucleic Acids Res. 16:1206-1206(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03831; CAA27446.1; -.
DR   PIR; A23386; SAZQGM.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1726       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    133    133       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    567    567       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    638    638       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    924    924       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    944    944       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    990    990       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1016   1016       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1114   1114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1221   1221       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1613   1613       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1726 AA;  196197 MW;  DD8AD45FA352BCF3 CRC64;


  Query Match             58.7%;  Score 1164;  DB 1;  Length 1726;
  Best Local Similarity   57.5%;  Pred. No. 4.7e-55;
  Matches  225;  Conservative   60;  Mismatches   84;  Indels   22;  Gaps    3;

Qy       7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db    1338 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 1397

Qy      67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db    1398 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 1457

Qy     127 SEKYKSDLDSIKKYI-----------------NDKQGENEKYLPFLNNIETLYKTVN 166
             ||| ||:|||| |                  |:| :  |:|||| ||||||  :
Db    1458 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 1517

Qy     167 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 226
           :||| ::|:|:|:||: :  ||    ||| :|: || | ||:  || :|:|  |  |  |
Db    1518 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 1577

Qy     227 NHNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLD 286
             ::| | ||||:| :|   :::| |:: |  : : : :     |    ::|||||||
Db    1578 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLD 1635

Qy     287 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDC 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1636 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDS 1695

Qy     347 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 377
           |||||||||||||||||||||||||||||||
Db    1696 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   4
MSP1_PLAFP
ID   MSP1_PLAFP     STANDARD;      PRT;  1726 AA.
AC   P50495;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (GP195).
GN   MSP-1.
OS   Plasmodium falciparum (isolate Palo Alto / Uganda).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=57270;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89005525; PubMed=3049134;
RA   Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
RA   Siddiqui W.A.;
RT   "Plasmodium falciparum: gene structure and hydropathy profile of the
RT   major merozoite surface antigen (gp195) of the Uganda-Palo Alto
RT   isolate.";
RL   Exp. Parasitol. 67:1-11(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M37213; AAA29611.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1726       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    133    133       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    567    567       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    638    638       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    924    924       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    944    944       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    990    990       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1016   1016       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1114   1114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1221   1221       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD   1613   1613       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1726 AA;  196174 MW;  5B59CEEFA2F9A026 CRC64;


  Query Match             58.7%;  Score 1163;  DB 1;  Length 1726;
  Best Local Similarity   57.5%;  Pred. No. 5.3e-55;
  Matches  225;  Conservative   60;  Mismatches   84;  Indels   22;  Gaps    3;

Qy       7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db    1338 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 1397

Qy      67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
           ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db    1398 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 1457

Qy     127 SEKYKSDLDSIKKYI-------------------NDKQGENEKYLPFLNNIETLYKTVN 166
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db    1458 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 1517

Qy     167 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 226
           :||| ::|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db    1518 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFDAIKKLINDD 1577

Qy     227 NHNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLD 286
             ::| | ||||:| :|   :::| |:: |  : : : :     |    ::|||||||
Db    1578 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLD 1635

Qy     287 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDC 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1636 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDS 1695

Qy     347 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 377
           |||||||||||||||||||||||||||||||
Db    1696 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   5
MSP1_PLAF3
ID   MSP1_PLAF3     STANDARD;      PRT;  1682 AA.
AC   P19598; Q25921;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P190).
GN   MSP-1.
OS   Plasmodium falciparum (isolate ro-33 / Ghana).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5834;
RN   [1]
RP   SEQUENCE OF 1-1061 FROM N.A.
RX   MEDLINE=88166657; PubMed=3327688;
RA   Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT   "A naturally occurring gene encoding the major surface antigen
RT   precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
RL   EMBO J. 6:4137-4142(1987).
RN   [2]
RP   SEQUENCE OF 1032-1682 FROM N.A.
RX   MEDLINE=95354793; PubMed=7628566;
RA   Tolle R., Bujard H., Cooper J.A.;
RT   "Plasmodium falciparum: variations within the C-terminal region of
RT   merozoite surface antigen-1.";
RL   Exp. Parasitol. 81:47-54(1995).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M35727; AAA29715.1; -.
DR   EMBL; Y00087; CAA68280.1; -.
DR   EMBL; Z35326; CAA84555.1; -.
DR   PIR; S06286; S06286.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1682       MEROZOITE SURFACE PROTEIN 1.
FT   TRANSMEM   1666   1682       MEMBRANE ANCHOR (POTENTIAL).
FT   CARBOHYD    233    233       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    462    462       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    528    528       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    599    599       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    785    785       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    881    881       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    901    901       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    947    947       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1071   1071       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1178   1178       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1569   1569       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1682 AA;  192462 MW;  C82A1E159948CAD6 CRC64;


  Query Match             58.4%;  Score 1156.5;  DB 1;  Length 1682;
  Best Local Similarity   57.7%;  Pred. No. 1.1e-54;
  Matches  225;  Conservative   58;  Mismatches   86;  Indels   21;  Gaps    3;

Qy       7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db    1295 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 1354

Qy      67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
           ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db    1355 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 1414

Qy     127 SEKYKSDLDSIKKYI-------------------NDKQGENEKYLPFLNNIETLYKTVND 167
             ||| ||:|||| |                    |:| :  |:|||| ||||||  : :
Db    1415 LAKYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVN 1474

Qy     168 KIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYN 227
           ||| ::|:|:||: :   ||    ||| :|: || | ||:  ||   :|  |  |  |
Db    1475 KIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFEAIKKLINDDT 1534

Qy     228 HNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDE 287
             ::| | ||||:| :|   :::| |:: |  : : : :     |    |:|||||||||
Db    1535 KKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPQNSGCFRHLDE 1592

Qy     288 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCY 347
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    1593 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 1652

Qy     348 PLFDGIFCSSSNFLGISFLLILMLILYSFI 377
           |||||||||||||||||||||||||||||||
Db    1653 PLFDGIFCSSSNFLGISFLLILMLILYSFI 1682


RESULT   6
MSP1_PLAFF
ID   MSP1_PLAFF     STANDARD;      PRT;  1701 AA.
AC   P13819;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA).
GN   MSP-1.
```

```
OS   Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88142999; PubMed=2449612;
RA   Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
RA   Brown G.V., Anders R.F., Kemp D.J.;
RT   "Variation in the precursor to the major merozoite surface antigens
RT   of Plasmodium falciparum.";
RL   Mol. Biochem. Parasitol. 27:291-302(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19143; AAA29653.1; -.
DR   PIR; A54498; A54498.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1701       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    536    536       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    607    607       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    802    802       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    899    899       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    965    965       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    991    991       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1089   1089       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1196   1196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1588   1588       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1701 AA;  193719 MW;  3920B75E73D38552 CRC64;


  Query Match             57.8%;  Score 1145;  DB 1;  Length 1701;
  Best Local Similarity   56.8%;  Pred. No. 4.7e-54;
  Matches  222;  Conservative   61;  Mismatches   86;  Indels   22;  Gaps    3;

Qy       7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
           :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db    1313 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 1372

Qy      67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db    1373 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 1432

Qy     127 SEKYKSDLDSIKKYI-------------------NDKQGENEKYLPFLNNIETLYKTVN 166
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db    1433 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 1492

Qy     167 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 226
           :||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db    1493 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 1552

Qy     227 NHNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLD 286
             ::| | ||||:| :|   :::| |:: |  : : : :     |    ::||||||||
Db    1553 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLD 1610

Qy     287 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDC 346
           |||||||||||||||||||||||||||||||||||||||| ||||||||: ||||||||||
Db    1611 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDS 1670

Qy     347 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 377
           |||||||||||||||||||||||||||||||
Db    1671 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 1701


RESULT   7
MSP1_PLAFM
ID   MSP1_PLAFM     STANDARD;      PRT;  1701 AA.
AC   P08569;
DT   01-AUG-1988 (Rel. 08, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P190).
GN   MSP-1.
OS   Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=70153;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88011243; PubMed=3079521;
RA   Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT   "Allelic dimorphism in a surface antigen gene of the malaria parasite
RT   Plasmodium falciparum.";
RL   J. Mol. Biol. 195:273-287(1987).
RN   [2]
RP   REVISIONS TO 1403; 1569 AND 1629.
RA   Tanabe K.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-115 FROM N.A.
RX   MEDLINE=86136024; PubMed=3004972;
RA   Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA   Stunnenberg H., Bujard H.;
RT   "Polymorphism of the precursor for the major surface antigens of
RT   Plasmodium falciparum merozoites: studies at the genetic level.";
RL   EMBO J. 4:3823-3829(1985).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X05624; CAA29112.1; -.
DR   PIR; A26868; A26868.
DR   PIR; B25120; B25120.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1701       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    536    536       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    607    607       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    802    802       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    899    899       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    965    965       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    991    991       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1089   1089       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1196   1196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1588   1588       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1701 AA;  193768 MW;  3FC2EC59AF96EA98 CRC64;
```

Query Match 57.8%; Score 1145; DB 1; Length 1701;
Best Local Similarity 56.8%; Pred. No. 4.7e-54;
Matches 222; Conservative 61; Mismatches 86; Indels 22; Gaps 3;

```
Qy      7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
          :|||||  ||||:|:||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db   1313 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 1372

Qy     67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
          ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db   1373 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 1432

Qy    127 SEKYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVN 166
            ||| ||:|||| |                  |:| :  |:|||| ||||||  :
Db   1433 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 1492

Qy    167 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 226
          :||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db   1493 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 1552

Qy    227 NHNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLD 286
             ::| | ||||:| :|   :::| |:: |  : : : :     |    ::|||||||
Db   1553 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLD 1610

Qy    287 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDC 346
          |||||||||||||||||||||||||||||||||||| ||||||||: ||||||||||
Db   1611 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDS 1670

Qy    347 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 377
          |||||||||||||||||||||||||||||||
Db   1671 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 1701
```

```
RESULT   8
MSP1_PLAYO
ID   MSP1_PLAYO     STANDARD;      PRT;  1772 AA.
AC   P13828;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (230 KDA).
GN   MSP-1.
OS   Plasmodium berghei yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5862;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90205979; PubMed=2320061;
RA   Lewis A.P.;
RT   "Sequence analysis upstream of the gene encoding the precursor to the
RT   major merozoite surface antigens of Plasmodium yoelii.";
RL   Mol. Biochem. Parasitol. 39:285-288(1990).
RN   [2]
RP   SEQUENCE OF 1093-1772 FROM N.A.
RC   STRAIN=17XL;
RX   MEDLINE=88124889; PubMed=2448778;
RA   Burns J.M. Jr., Daly T.M., Vaidya A.B., Long C.A.;
RT   "The 3' portion of the gene for a Plasmodium yoelii merozoite surface
RT   antigen encodes the epitope recognized by a protective monoclonal
RT   antibody.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:602-606(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03612; AAA29762.1; -.
DR   EMBL; J04668; AAA29702.1; -.
DR   PIR; A28121; A28121.
DR   PIR; A45532; A45532.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   1772       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    406    406       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    646    646       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    829    829       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1018   1018       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1090   1090       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1408   1408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1446   1446       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1541   1541       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1629   1629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1680   1680       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1521   1521       L -> V (IN REF. 2).
SQ   SEQUENCE   1772 AA;  197230 MW;  9A6291658EB0F45D CRC64;
```

Query Match 29.7%; Score 588.5; DB 1; Length 1772;
Best Local Similarity 34.4%; Pred. No. 2.4e-24;
Matches 132; Conservative 79; Mismatches 142; Indels 31; Gaps 7;

```
Qy      9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVLES 68
          :|||:  ||   :| | |   |:|||| :     |   :: : ||::  || :|  ||
Db   1401 DILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQKRNDFLEVLSH 1460

Qy     69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
          :|  :|||::: ||:::||: |: :|:|| : :  |    |: ||    | : ::  : |
Db   1461 ELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFNKMVE 1520

Qy    129 KYKSDLDSIKKYI-------ND-KQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
            | : | ::|: |       ||  : | :||:| | ::: ||:||  : : :   |: ::
Db   1521 LYNTQLAAVKEQIATIEAETNDTNKEEKKKYIPILEDLKGLYETVIGQAEEYSEELQNRL 1580

Qy    181 LNYTYEKSNVEVKIKELNYLKTIQDKLADF----KKNNNFVGIADLSTDYNHNNLLTKFL 236
          ||  ||:  |:  | |     | :|| :|    : | :   ||           |
Db   1581 DNYKNEKAEFEILTKNLEKYIQIDEKLDEFVEHAENNKHIASIA----------LNNLN 1629

Qy    237 STGMVFENLLKSVLSNLLDWK----LARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECK 292
           :|:| |   | :|: :|:      |    ||      |    | :::|||  :  || :
Db   1630 KSGLVGEGESKKILAKMLNMDGMDLLGVDPKHVCVDTRD---IPKNAGCFRDDNGTEEWR 1686

Qy    293 CLLNYKQ-EGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFD 351
          ||| ||: ||: |||| |||| |||||  | |   :|  | ||| | | :|     ::
Db   1687 CLLGYKKGEGNTCVENNNPTCDINNGGCDPTASCQNAESTENSKKIICTCKEPTPNAYYE 1746

Qy    352 GIFCSSSNFLGISFLLILMLILYS 375
          |:|||||:|:|:| |||: ||:::
Db   1747 GVFCSSSSFMGLSILLIITLIVFN 1770
```

```
RESULT   9
RBP2_PLAVB
ID   RBP2_PLAVB     STANDARD;      PRT;  1251 AA.
AC   Q00799;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
```

```
DE   RETICULOCYTE BINDING PROTEIN 2 (FRAGMENT).
GN   RBP2.
OS   Plasmodium vivax (strain Belem).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=31273;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92315338; PubMed=1617731;
RA   Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT   "A reticulocyte-binding protein complex of Plasmodium vivax
RT   merozoites.";
RL   Cell 69:1213-1226(1992).
CC   -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC       HUMAN RETICULOCYTE CELLS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND (PROBABLE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M88098; AAA29744.1; -.
KW   Malaria; Receptor; Membrane.
FT   NON_TER       1      1
FT   NON_TER    1251   1251
SQ   SEQUENCE   1251 AA;  143741 MW;  54BA51C7404AC572 CRC64;


  Query Match             7.2%;  Score 142;  DB 1;  Length 1251;
  Best Local Similarity   23.2%;  Pred. No. 1.1;
  Matches   78;  Conservative   53;  Mismatches  117;  Indels   88;  Gaps   15;

Qy       3 TTSVIDNILSKIENEYEVL--YLKPLAGVYRSLKKQLEN--NVMTFNVNVKDILNSRFNK 58
Db     891 TNTNSDSKLEKIKTDFESLKNALKTLEGEVNALKASSDNHEHVQSKSEPVNPAL-SEIEK 949

Qy      59 RENFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSI------DTD 112
Db     950 EET------------DIDSLNTALDE----LLKKGRTCEVSRYKLIKDTVTKEISDDTE 992

Qy     113 I--NFANDVLGYYKILSEKYK---SDLDSIKKYINDKQGENEKYLPF------------ 154
Db     993 LINTIEKNVKAYLAYIKKNYEDTVQDVLTLNEHFNTKQVSNHEPTNFDKSNKSSEELTKA 1052

Qy     155 -------LNNIETLYKTVNDKIDLFVIHLEAK-------------VLNYTYEKSNVEVK 193
Db    1053 VTDSKTIISKLKGVIIEVNENTEMNTIESSAKEIEALYNELKNKKTSLNEIYQTSN-EVK 1111

Qy     194 IKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGMVFENLLKSVLSNL 253
Db    1112 LQE---MKSNADKYIDVSKIFNTV------LDTQKSNIVTNQHSINNV-KDKLKGKLQEL 1161

Qy     254 LD----------WKLARYVKHFTTPMRKKTMIQQSS 279
Db    1162 IDADSSFTLESIKKFNEIYSHIKTNIGELEQLQQTN 1197


RESULT  10
Y704_METJA
ID   Y704_METJA     STANDARD;      PRT;   377 AA.
AC   Q58115;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ0704.
GN   MJ0704.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67517; AAB98705.1; -.
DR   TIGR; MJ0704; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     26     46       POTENTIAL.
FT   TRANSMEM     67     87       POTENTIAL.
FT   TRANSMEM    108    128       POTENTIAL.
FT   TRANSMEM    135    155       POTENTIAL.
SQ   SEQUENCE   377 AA;  44630 MW;  2746A0A96425C285 CRC64;


  Query Match             7.0%;  Score 139.5;  DB 1;  Length 377;
  Best Local Similarity   25.0%;  Pred. No. 0.39;
  Matches   57;  Conservative   39;  Mismatches   63;  Indels   69;  Gaps   12;

Qy      26 LAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVLESDLIPYKDLTSSNYVVKD 85
Db     194 LFAVYEIIEKRIKVGDVYAVIKCLNMINKNFNKY------------------------- 227

Qy      86 PYKFLNKEKRDKFLSSYNYIKDSIDTD-----INFAND--------------VLGYYKI 125
Db     228 -WMFIKDEKREKYLRDFLRILSKLRVEYRKNCLNRKNKPYSKKGLEAFKKTEFIVSFYKT 286

Qy     126 LSEKYKS-DLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDK--IDLFVIHLEAKVLN 182
Db     287 IEEKIKTRDINSINKFL------SETY----NNISN-YEMFNDKTYLEIFVHHL--KELL 333

Qy     183 YTYEKSNVEVKIKE--LNYLKTIQDKLADFKKNNNFVGIADLSTDYNH 228
Db     334 EKYEKEKNENKLDEEIFNELKDELEKLIN-KCNNK---LRELESQNNN 377


RESULT  11
YLJ2_CAEEL
ID   YLJ2_CAEEL     STANDARD;      PRT;   2198 AA.
AC   P34367;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 256.3 KDA PROTEIN C50C3.2 IN CHROMOSOME III.
GN   C50C3.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
```

```
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L14433; AAA27973.1; -.
DR   WormPep; C50C3.2; CE01861.
DR   InterPro; IPR002017; -.
DR   InterPro; IPR002048; -.
DR   Pfam; PF00435; spectrin; 12.
DR   PROSITE; PS00018; EF_HAND; 1.
KW   Hypothetical protein; Calcium-binding.
FT   CA_BIND    2025   2036       POTENTIAL.
SQ   SEQUENCE   2198 AA;  256260 MW;  CD1B2C1092C5EDC8 CRC64;
```

```
  Query Match             6.9%;  Score 136.5;  DB 1;  Length 2198;
  Best Local Similarity   21.3%;  Pred. No. 4;
  Matches   69;  Conservative   66;  Mismatches   98;  Indels   91;  Gaps   15;

Qy       8 DNILSKIENEYEVLYLKPLAGVYRSLKKQLE--------NNVMTFNVNVKDILNSRFNKR 59
           :|:||: |   ::| ||     :| |:|::|        :::: : : :|:   |  : :
Db      72 NNVLSRWEYLLKLLALK-----WRQLEKEIESIEFKNKCDSIVDWILKIKE---SDESGK 123

Qy      60 ENFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDV 119
           | |  : :  :: : :::::   : ||                                | |
Db     124 EIFDEIRKCSMV-WNEISATFQQMLDP----------------------------TATDT 154

Qy     120 LGYYKILS--EKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLE 177
             | |:      ::  :||: |    |  |||::  |::| | | | ::||
Db     155 SNYEKVHETWSNFQEYIDSLHK----KLSENERFQKFVDNAEDLIKWMDDK--------- 201

Qy     178 AKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLS 237
           | :   | | : || :|   | ||:: ::   ::      | ||   :
Db     202 EKEICEKYSKMDFEVMMK---YRKTVEIEMKSVEQR-----IKDLKEQF----------- 242

Qy     238 TGMVFENLLKSVLSNLLDWK-----LARYVKHFTTPMRK-KTMIQQSSGCFRHLDEREEC 291
           ||  :|     | |: | |     : :  : |   ::| || |: |:   : : | |
Db     243 VGMENDN-----LRNIPDPKNHVIDVEQRFESFQAFVQKWKTDIENSADADKLMKEAENI 297

Qy     292 KCLLNYKQEGDKCV-ENPNPTCNE 314
           |  : | |  | :  :  | |:|
Db     298 CCWSSEKIEDLKIMATSDTPDCDE 321
```

```
RESULT  12
RPOP_AGABT
ID   RPOP_AGABT     STANDARD;      PRT;  1102 AA.
AC   P33539;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PROBABLE DNA-DIRECTED RNA POLYMERASE (EC 2.7.7.6).
OS   Agaricus bitorquis.
OG   Mitochondrion.
OG   Plasmid pEM.
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Agaricales;
OC   Agaricaceae; Agaricus.
OX   NCBI_TaxID=5343;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 24666 / AG4;
RX   MEDLINE=91347410; PubMed=1879001;
RA   Robison M.M., Royer J.C., Horgen P.A.;
RT   "Homology between mitochondrial DNA of Agaricus bisporus and an
RT   internal portion of a linear mitochondrial plasmid of Agaricus
RT   bitorquis.";
RL   Curr. Genet. 19:495-502(1991).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC       RNA(N).
CC   -!- SIMILARITY: BELONGS TO THE PHAGE AND MITOCHONDRIAL RNA POLYMERASES
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X63075; CAA44799.1; -.
DR   PIR; S28104; S28104.
DR   InterPro; IPR002092; -.
DR   Pfam; PF00940; RNA_pol; 1.
DR   PROSITE; PS00489; RNA_POL_PHAGE_2; 1.
DR   PROSITE; PS00900; RNA_POL_PHAGE_1; 1.
KW   Transferase; Transcription; DNA-directed RNA polymerase;
KW   Mitochondrion; Plasmid.
FT   ACT_SITE    734    734       BY SIMILARITY.
FT   ACT_SITE    804    804       BY SIMILARITY.
FT   ACT_SITE    980    980       BY SIMILARITY.
SQ   SEQUENCE   1102 AA;  126664 MW;  F72FDD308D1AB5B4 CRC64;
```

```
  Query Match             6.8%;  Score 135;  DB 1;  Length 1102;
  Best Local Similarity   22.5%;  Pred. No. 2.2;
  Matches   73;  Conservative   57;  Mismatches  131;  Indels   64;  Gaps   14;

Qy       5 SVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE-NFK 63
           |::||:   : :              | |   | ::|::  :  |: :||   |  : |
Db      27 SLVDNLFYNLHD-~~~~~~~~~~~YAS---QSKHNIIIDDGKVEKLLNLICNVLDRNVN 70

Qy      64 NVLESDLIPYKDLTS--SNYVVKDPYKFLNKEK---RDKFLSSYNYIKD---SIDTD-IN 114
             : :|  :  ||:      ||    |: :   |   |:   ||| | |:    :|   |:
Db      71 ELNDSVFLIIKDIEKECKYYVSNVLYRSVGSRKNRGREVEWSSYKYNKEFNKVLDKGIIS 130

Qy     115 FANDVLGYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVI 174
             |:|| :     | |   ::||   : :       | |   |||  :  :: :|:
Db     131 INNEVLKFISKEREGYIERVESIAVTVKN------KILELNNNIAEVLLSIKNKV----- 179

Qy     175 HLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNF-----VGIADLSTDYNHN 229
                 |||    | :|  |::|:||   : :|      | ||       |  :  : |
Db     180 ----IVLN----KESVVAKVEEINY--EVHNKFIKGNGNTNFSNRNLTEIKSILKELNKM 229

Qy     230 NLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDERE 289
           :|   ::     |: |  |:  :||  |    |       :||:::     |||  :
Db     230 EILDNRINKLSTKESDLLKVIKEILDSNLIIEDKQLAI---EKTVVEYELTFFRHNMDTH 286

Qy     290 ECKCLLNYKQEGDKCVENPNPTCNE 314
           | :       :| : |  |  |:
Db     287 ETR---------NKIIHNIYPKLNK 302
```

```
RESULT 13
IRR1_YEAST
ID   IRR1_YEAST     STANDARD;      PRT;  1150 AA.
AC   P40541;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   IRR1 PROTEIN.
GN   IRR1 OR YIL026C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=96090137; PubMed=7483852;
RA   Kurlandzka A., Rytka J., Gromadka R., Murawski M.;
RT   "A new essential gene located on Saccharomyces cerevisiae chromosome
RT   IX.";
RL   Yeast 11:885-890(1995).
CC   -!- FUNCTION: NOT KNOWN; MAY PARTICIPATE IN CELL-CELL AND/OR CELL-
CC       SUBSTRATUM INTERACTIONS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: SOME, TO S.POMBE SPAC17H9.20.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z46881; CAA86966.1; -.
DR   SGD; S0001288; IRR1.
FT   DOMAIN       44     48       POLY-GLU.
FT   DOMAIN       65     70       POLY-ASP.
SQ   SEQUENCE   1150 AA;  133008 MW;  89688EA09485AC28 CRC64;

  Query Match             6.8%;  Score 134;  DB 1;  Length 1150;
  Best Local Similarity   20.6%;  Pred. No. 2.7;
  Matches   89;  Conservative   74;  Mismatches  134;  Indels  136;  Gaps   22;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLE----NNVMTFNVNVKDILNSRF 56
           |   | | |  ||:: :|:: : | :  ::   ||:|:    | : |  :::|  |  :
Db     733 AFMDSTIGN--SKVDMKYDI-FAKFIHHIHHFEKKELQEKFLNQIATLKIHLKKFLQEKM 789

Qy      57 ---NKRENFKNVLESDLIPYKDLTSSNYVVKDPYKFLNK----------EKRDKFLSSY- 102
              | |::           |||| | | :     ::||          |  :: | :
Db     790 DPNNSRDD-----------YKDLTCSLYEL-----YINKLTILGRDYPIEVDEELLQLFL 833

Qy     103 -NYIK----------DSIDTDINFANDVLGYYKILS----EKYKSDLDSIKKYIN----D 143
            |::          ||   :|||   ||     |:    ||::  :: :: | |    |
Db     834 NNFVSRIPIMFQDFDDSTAQEINFKMLVL-----LATWNLEKWREIIEKVRDYENSISKD 888

Qy     144 KQGENEKYLPFLNNIETL----------YKTVND----------------KIDLFVIHL 176
           :   :     :  :||          :: :|                  | :| : |
Db     889 LRSVWKPIAAIIGRLNTLVISLAATNETFENINSLFYLKWSACTSLMDIIVAIKIFELKL 948

Qy     177 EAKVLNYTYEKSN----------VEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 226
           |   : |  |           :| :|   ||:::  |  | :      :||   :
Db     949 PADATTWRYSMSEQFPFYLHDNASKVLLKIFLYLESLFAKQVDVQLER----VADEDANL 1004

Qy     227 N-------HNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYV---KHFTTPMRKKTM-- 274
           |         |: |:||   :  : |:|    |:|| : |  |   |    |: || :
Db    1005 NDLPETGFFENIETEFLLFTVKLKGLMK---LNILDERFASRVALNKEKLGPLFKKIVDD 1061

Qy     275 ------------IQQSSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDAD 322
                       ||::      :  :||:     | ::| |    : |   |: :     |
Db    1062 TIMENPEPNKKNIQKAKS---NQTQREKAPLQPNSERETD----HANTENNDPDIPMTID 1114

Qy     323 AKCTEEDSGSNGK 335
           :  || | :| :
Db    1115 LEPIEESSQNNSE 1127


RESULT 14
Y875_METJA
ID   Y875_METJA     STANDARD;      PRT;   748 AA.
AC   Q58285;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ0875.
GN   MJ0875.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67531; AAB98882.1; -.
DR   HSSP; P08622; 1XBL.
DR   TIGR; MJ0875; -.
KW   Hypothetical protein.
SQ   SEQUENCE   748 AA;  90731 MW;  5515ED3B541562EA CRC64;

  Query Match             6.7%;  Score 133;  DB 1;  Length 748;
  Best Local Similarity   23.0%;  Pred. No. 1.9;
  Matches   72;  Conservative   38;  Mismatches  101;  Indels  102;  Gaps   15;

Qy      17 EYEVLYLKPLAGVYRSLK--------------KQLENNVMTFN-VNVK----------D 50
           : :: | | |:  :|  :                |  | :   ||  | |           :
Db     154 DIKIYYYKKLSIKHRKFEGNLDKAIELTEKCYKLAEESYNKFNDKNYKKAEIFNKHFYYN 213

Qy      51 ILNSRFNKRENFKNVLE-----SDLI--------------PYKDLTSSNYVVKDPY-KFL 90
           ::  :|     ||   |       | |               || |   |   |: : :::
Db     214 LMAQKFESERKFKEAAEYYKKSGDTIKEIDEKIAYDEYANSYKWLAIENKYNKEKFEEYI 273

Qy      91 NK-----EKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSEKYKSDLDSIKK----YI 141
           ||     |||   |  | |:    |  : ||||       | ||    :| |||    |
Db     274 NKAIEFSEKRGDKLQEYYYLGLKYDHLVREAND-------LEEK----IDYIKKSKEYYY 322
```

```
Qy     142 NDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLK 201
           |  |  ||: :|                :    : ::||  |||:        ||:|:
Db     323 RSKSFEFAKYMEYLE--------------YYYQFKYELLNGNYEKA--------LNFLQ 359

Qy     202 TIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKF---LSTG------MVFENLLKSVLSN 252
           :  | : | :|      |  |     |: :|   :| |       : :  |:  |
Db     360 KAKKSLKNVKISNIIFSKYTLECD----ELICRFYLSISQGEFKKSVELLDEYLEISLKI 415

Qy     253 LLDWKLARYVKHF 265
           | |||  |  | :
Db     416 LSDWKNTRKYKFY 428


RESULT 15
CHEA_BORBU
ID   CHEA_BORBU     STANDARD;      PRT;   864 AA.
AC   Q44737; P70857; Q44877;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CHEMOTAXIS PROTEIN CHEA (EC 2.7.3.-).
GN   CHEA OR BB0669.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=212;
RX   MEDLINE=98438936; PubMed=9765799;
RA   Trueba G.A., Old I.G., Saint-Girons I., Johnson R.C.;
RT   "A cheA cheW operon in Borrelia burgdorferi, the agent of Lyme
RT   disease.";
RL   Res. Microbiol. 148:191-200(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
RN   [3]
RP   SEQUENCE OF 1-30 FROM N.A.
RC   STRAIN=212;
RX   MEDLINE=97144545; PubMed=8990312;
RA   Ge Y., Charon N.W.;
RT   "An unexpected flaA homolog is present and expressed in Borrelia
RT   burgdorferi.";
RL   J. Bacteriol. 179:552-556(1997).
CC   -!- FUNCTION: INVOLVED IN THE TRANSMISSION OF SENSORY SIGNALS FROM
CC       THE CHEMORECEPTORS TO THE FLAGELLAR MOTORS. CHEA IS
CC       AUTOPHOSPHORYLATED; IT CAN TRANSFER ITS PHOSPHATE GROUP TO EITHER
CC       CHEB OR CHEY (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC   -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC       KINASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U28962; AAB96835.1; -.
DR   EMBL; AE001168; AAC67024.1; -.
DR   EMBL; U62900; AAC44771.1; -.
DR   EMBL; X91907; CAA63002.1; -.
DR   TIGR; BB0669; -.
DR   InterPro; IPR000410; -.
DR   InterPro; IPR002545; -.
DR   InterPro; IPR002570; -.
DR   Pfam; PF01584; CheW; 1.
DR   Pfam; PF01627; Hpt; 1.
DR   Pfam; PF00512; signal; 1.
DR   PRINTS; PR00344; BCTRLSENSOR.
KW   Sensory transduction; Transferase; Kinase; Phosphorylation;
KW   Chemotaxis.
FT   MOD_RES      51     51       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CONFLICT     11     11       L -> I (IN REF. 1).
FT   CONFLICT    182    182       S -> H (IN REF. 1).
FT   CONFLICT    187    187       S -> G (IN REF. 1).
FT   CONFLICT    239    239       L -> S (IN REF. 1).
FT   CONFLICT    362    362       A -> S (IN REF. 1).
FT   CONFLICT    559    559       S -> P (IN REF. 1).
SQ   SEQUENCE   864 AA;  98352 MW;  C1111DE0877BE624 CRC64;


  Query Match             6.7%;  Score 133;  DB 1;  Length 864;
  Best Local Similarity   20.3%;  Pred. No. 2.2;
  Matches   84;  Conservative   64;  Mismatches  121;  Indels  144;  Gaps   19;

Qy       8 DNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVLE 67
             |  | ::  :| ::|   : || :||        : |: ::  |   || :|  | : :
Db     214 DKFLKRV--DYYLIY-SDIEGVKKSLD--------SLNL-IESYLVDEFNVKEELKKLAD 261

Qy      68 SDLIPYKDL-TSSNYVVKDPYKFLNKEKRDKFLSSYN---------YIKDSIDTDINFAN 117
             ::   ||:   ||:|: | : |   |  |  |   |         ::||:    |:
Db     262 EEI---KDVDLDSNFVLNDNFDFTEDEISDLLLEVENQKLFKVRLDFVKDNPMATIS-GL 317

Qy     118 DVLGYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDL--FVIH 175
           :|   | | : :||  ||  : : ||  :   |    |  |   ::: ||:|   | |
Db     318 QMLQALKSLGKIFKSIPDS-SELLADKFFDFVIYYLISNTSE---ESIAKKINLPDVVSH 373

Qy     176 LEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTD---------- 225
           | | :|    ||   |::||        |  | ||:| |    : :| :
Db     374 FEIKNVNLESLKS---VRLKE--------DDEAPFKENKNIKKNSPISVNLIRIDSKKID 422

Qy     226 ----------------------------YNHN--------------NLLTKFLSTGMVFE 243
                                       || |              :|  |  |:  |
Db     423 YILNLVSEAVISKSSYNQINSEMITLFYNFNYFYDYQESFQRNFLIDLKIVFKDAGLTLE 482

Qy     244 NLLKSVLSNLLDWKLARYVKHFT----------------------TPMRK---KTMIQ 276
           : ::| :::|: :|:. : :|  :                      | : :   || :
Db     483 DEIESHINSLMSFKMEKALKDISELRNSFFRLLQNFKMTSGRLSRIITDLHESVLKTRML 542

Qy     277 QSSGCFRHL-----DEREECKCLLNYKQEGDK----------------CVEN 307
             |  |        |  ::   ::| | ||::                 || |
Db     543 PISNIFSRFTRVVRDLSKKLNKIVNLKMEGEETELDKSVIDDLVDPLMHCVRN 595


Search completed: August  8, 2001, 12:32:42
Job time: 210 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:30:58 ; Search time 43.63 Seconds
(without alignments)
658.212 Million cell updates/sec

Title: US-09-500-376-5
Perfect score: 1982
Sequence: 1 AVTTSVIDNILSKIENEYEV.....SNFLGISFLLILMLILYSFI 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1947	98.2	1631	1	SAZQK1	major merozoite su
2	1942	98.0	1640	2	A24594	probable major sur
3	1824.5	92.1	400	2	A45545	major merozoite su
4	1824.5	92.1	1639	2	S05603	major merozoite su
5	1167.5	58.9	651	2	S47282	merozoite surface
6	1164	58.7	1726	1	SAZQGM	major merozoite su
7	1157	58.4	1726	2	A45948	major merozoite su
8	1145	57.8	1701	2	A54498	major merozoite su
9	1131	57.1	1701	2	A26868	major merozoite su
10	745	37.6	1726	2	A39401	merozoite surface
11	744	37.5	1751	2	A45604	major blood-stage
12	588.5	29.7	680	2	A28121	major merozoite su
13	588.5	29.7	1772	2	A45532	major merozoite su
14	546	27.5	1785	2	A45546	major merozoite su
15	157.5	7.9	3394	2	T18501	hypothetical prote
16	155.5	7.8	2269	2	T28677	rhoptry protein -
17	155.5	7.8	3973	2	B71612	hypothetical prote
18	154	7.8	2010	2	B71616	phosphatase (acid
19	153.5	7.7	1306	2	T28313	ORF MSV152 probabl
20	152.5	7.7	2401	2	T28676	rhoptry protein -
21	149	7.5	507	2	A71622	hypothetical prote
22	149	7.5	722	2	D71607	VPS45-like protein
23	149	7.5	980	2	E71606	hypothetical prote
24	149	7.5	1346	2	G71613	hypothetical prote
25	148.5	7.5	706	2	A57719	dihydropteroate sy
26	148.5	7.5	1619	2	T18499	hypothetical prote
27	147.5	7.4	1127	2	T28317	ORF MSV156 hypothe
28	145	7.3	1714	2	E71609	Ser/Thr protein ki
29	144.5	7.3	1119	2	T18491	hypothetical prote
30	142	7.2	481	2	T18465	hypothetical prote
31	142	7.2	1252	2	B42771	reticulocyte-bindi
32	142	7.2	1939	2	T18372	repeat organellar
33	142	7.2	3844	2	T18402	asparagine/asparta
34	141	7.1	608	2	T28301	ORF MSV140 hypothe
35	141	7.1	1005	2	A64465	hypothetical prote
36	141	7.1	1187	2	T18355	hypothetical prote
37	141	7.1	2500	2	G71609	hypothetical prote
38	140.5	7.1	2708	2	T09079	probable chloroqui
39	140.5	7.1	2819	2	T09080	probable chloroqui
40	140	7.1	1308	2	E71622	probable membrane
41	139.5	7.0	377	2	H64387	hypothetical prote
42	139	7.0	2523	2	T18477	hypothetical prote
43	138.5	7.0	2013	2	C71610	probable membrane
44	138.5	7.0	2166	2	G70163	hypothetical prote
45	138	7.0	821	2	S67087	hypothetical prote

ALIGNMENTS

RESULT 1
SAZQK1
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000
C;Accession: A25120
R;Mackay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.
EMBO J. 4, 3823-3829, 1985
A;Title: Polymorphism of the precursor for the major surface antigens of Plasmodium f
A;Reference number: A91030; MUID:86136024
A;Accession: A25120
A;Molecule type: DNA
A;Residues: 1-1631 <MAC>
C;Comment: The merozoite stages of different strains have strain-specific surface ant
C;Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane pr
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1631/Product: major merozoite surface antigen #status predicted <MAT>
F;67-84/Region: 3-residue repeats (S-G-T/P)
F;1614-1631/Domain: membrane anchor #status predicted <MBN>
F;97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (Asn

Query Match 98.2%; Score 1947; DB 1; Length 1631;
Best Local Similarity 98.4%; Pred. No. 2.8e-95;
Matches 371; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```
Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1255 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 1314

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1315 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 1374

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1375 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 1434

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    1435 LNYTYEKSNVEVKIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 1494

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
           ||||||||:|||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db    1495 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 1554

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
           | ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    1555 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 1614
```

```
Qy    361 LGISFLLILMLILYSFI 377
          |||||||||||||||||
Db   1615 LGISFLLILMLILYSFI 1631
```

RESULT 2
A24594
probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodium
C;Species: Plasmodium falciparum
C;Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jun-2000
C;Accession: A24594
R;Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls,
Nature 317, 270-273, 1985
A;Title: Primary structure of the precursor to the three major surface antigens of Plasm
A;Reference number: A24594; MUID:86014355
A;Accession: A24594
A;Molecule type: DNA
A;Residues: 1-1640 <HOL>
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

```
  Query Match             98.0%;  Score 1942;  DB 2;  Length 1640;
  Best Local Similarity   98.7%;  Pred. No. 5.2e-95;
  Matches  372;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1264 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 1323

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 1383

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1384 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 1443

Qy    181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1444 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 1503

Qy    241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1504 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 1563

Qy    301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
          | ||||| ||||||||||||||||||||||||||||||||||||||||   ||||||||
Db   1564 GSKCVENSNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLSMVIFCSSSNF 1623

Qy    361 LGISFLLILMLILYSFI 377
          |||||||||||||||||
Db   1624 LGISFLLILMLILYSFI 1640

RESULT 3
A45545
major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments)
C;Species: Plasmodium falciparum
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
C;Accession: A45545
R;Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.
Mol. Biochem. Parasitol. 49, 29-33, 1991
A;Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protein-1
A;Reference number: A45545; MUID:92131048
A;Accession: A45545
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-400 <BLA>
A;Note: sequence extracted from NCBI backbone (NCBIN:77612, NCBIP:77621)
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

```
  Query Match   92.1%; Score 1824.5; DB 2; Length 400;
  Best Local Similarity   93.6%; Pred. No. 1.7e-89;
  Matches 353; Conservative   6; Mismatches   17; Indels   1; Gaps   1;

Qy   1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 84

Qy   61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 144

Qy   121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   145 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 204

Qy   181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   205 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 264

Qy   241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
          ||||| |:||||||| | : : : : | |:||||||||||||||||||||||
Db   265 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 323

Qy   301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   324 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 383

Qy   361 LGISFLLILMLILYSFI 377
          |||||||||||||||||
Db   384 LGISFLLILMLILYSFI 400

RESULT 4
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N;Alternate names: gp195 surface antigen
C;Species: Plasmodium falciparum
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C;Accession: S05603; S04850
R;Myler, P.J.
submitted to the EMBL Data Library, April 1989
A;Reference number: S05603
A;Accession: S05603
A;Molecule type: mRNA
A;Residues: 1-1639 <MYL>
A;Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
R;Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989
A;Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Pl
A;Reference number: S04850; MUID:89345116
A;Accession: S04850
A;Molecule type: mRNA
A;Residues: 1504-1639 <MYL2>
A;Cross-references: EMBL:X15063
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1639/Product: major merozoite surface antigen #status predicted <MAT>

```
  Query Match             92.1%;  Score 1824.5;  DB 2;  Length 1639;
  Best Local Similarity   93.6%;  Pred. No. 8.2e-89;
  Matches  353;  Conservative    6;  Mismatches   17;  Indels    1;  Gaps    1;

Qy      1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1264 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 1323

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 1383
```

```
Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1384 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 1443

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1444 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 1503

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
           ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||||
Db    1504 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 1562

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    1563 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 1622

Qy     361 LGISFLLILMLILYSFI 377
           |||||||||||||||||
Db    1623 LGISFLLILMLILYSFI 1639


RESULT   5
S47282
merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71) (f
C;Species: Plasmodium falciparum
A;Variety: strain RO-71
C;Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 09-Jun-2000
C;Accession: S47282
R;Tolle, R.; Bujard, H.; Cooper, J.A.
submitted to the EMBL Data Library, July 1994
A;Description: Plasmodium falciparum: recombination within the C-terminal region of merc
A;Reference number: S47282
A;Accession: S47282
A;Molecule type: DNA
A;Residues: 1-651 <TOL>
A;Cross-references: EMBL:Z35329; NID:g535257; PIDN:CAA84558.1; PID:g535258
A;Experimental source: strain RO-71
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen


  Query Match             58.9%;  Score 1167.5;  DB 2;  Length 651;
  Best Local Similarity   57.9%;  Pred. No. 1.3e-54;
  Matches  226;  Conservative   59;  Mismatches   84;  Indels   21;  Gaps    3;

Qy       7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db     264 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 323

Qy      67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db     324 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 383

Qy     127 SEKYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVND 167
             ||| ||:|||| |                  |:| :  |:|||| ||||||  : :
Db     384 LAKYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVN 443

Qy     168 KIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYN 227
           ||| ::|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db     444 KIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDDT 503

Qy     228 HNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDE 287
           ::| | ||||:| :|   :::| |:: |  : : : :     |    |:|||||||
Db     504 KKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPQNSGCFRHLDE 561

Qy     288 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCY 347
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     562 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 621

Qy     348 PLFDGIFCSSSNFLGISFLLILMLILYSFI 377
           |||||||||||||||||||||||||||||
Db     622 PLFDGIFCSSSNFLGISFLLILMLILYSFI 651


RESULT   6
SAZQGM
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N;Alternate names: 195K glycoprotein
C;Species: Plasmodium falciparum
C;Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
C;Accession: A23386; S06361
R;Weber, J.L.; Leininger, W.M.; Lyon, J.A.
Nucleic Acids Res. 14, 3311-3323, 1986
A;Title: Variation in the gene encoding a major merozoite surface antigen of the huma
A;Reference number: A23386; MUID:86205236
A;Accession: A23386
A;Molecule type: DNA
A;Residues: 1-1104 <WEB1>
A;Cross-references: EMBL:X03831
R;Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.
Nucleic Acids Res. 16, 1206, 1988
A;Title: Merozoite surface protein sequence from the Camp strain of the human malaria
A;Reference number: S06361; MUID:88143999
A;Accession: S06361
A;Molecule type: DNA
A;Residues: 1104-1726 <WEB2>
A;Cross-references: EMBL:X03831
C;Comment: The merozoite stages of different strains have strain-specific surface ant
C;Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1726/Product: major merozoite surface antigen #status predicted <MAT>
F;67-87,91-96,100-105,109-120/Region: 3-residue repeats (S-G-T)
F;757-765/Region: 3-residue repeats (T-E-E)
F;133,272,501,567,638,827,839,924,944,990,1016,1114,1221,1613,1658/Binding site: carb


  Query Match             58.7%;  Score 1164;  DB 1;  Length 1726;
  Best Local Similarity   57.5%;  Pred. No. 5.9e-54;
  Matches  225;  Conservative   60;  Mismatches   84;  Indels   22;  Gaps    3;

Qy       7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db    1338 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 1397

Qy      67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db    1398 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 1457

Qy     127 SEKYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVN 166
             ||| ||:|||| |                  |:| :  |:|||| ||||||  :
Db    1458 LAKYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 1517

Qy     167 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 226
           :||| ::|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db    1518 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 1577

Qy     227 NHNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLD 286
           ::| | ||||:| :|   :::| |:: |  : : : :     |    ::|||||||
Db    1578 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLD 1635

Qy     287 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDC 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1636 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDS 1695

Qy     347 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 377
           |||||||||||||||||||||||||||||||
Db    1696 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   7
A45948
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
```

```
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jun-2000
C;Accession: A45948
R;Chang, S.P.; Kramer, K.J.; Yamaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.
Exp. Parasitol. 67, 1-11, 1988
A;Title: Plasmodium falciparum: gene structure and hydropathy profile of the major meroz
A;Reference number: A45948; MUID:89005525
A;Accession: A45948
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1726 <CHA>
A;Cross-references: GB:M37213
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match             58.4%;  Score 1157;  DB 2;  Length 1726;
  Best Local Similarity   57.3%;  Pred. No. 1.4e-53;
  Matches  224;  Conservative   60;  Mismatches   85;  Indels   22;  Gaps    3;

Qy       7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db    1338 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 1397

Qy      67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db    1398 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 1457

Qy     127 SEKYKSDLDSIKKYI--------------------NDKQGENEKYLPFLNNIETLYKTVN 166
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db    1458 LAKYKDDLESIKKVIKEEKEKFPSSPPTTTPSPVKTDEQKKESKFLPFLTNIETLYNNLV 1517

Qy     167 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 226
           :| | ::|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db    1518 NKNDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 1577

Qy     227 NHNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLD 286
             ::| | ||||:| :|   :::| |:: |  : : : :     |    ::||||||||
Db    1578 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLD 1635

Qy     287 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDC 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1636 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDS 1695

Qy     347 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 377
           |||||||||||||||||||||||||||||||
Db    1696 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   8
A54498
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (is
C;Species: Plasmodium falciparum
C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C;Accession: A54498
R;Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.;
Mol. Biochem. Parasitol. 27, 291-302, 1988
A;Title: Variation in the precursor to the major merozoite surface antigens of Plasmodiu
A;Reference number: A54498; MUID:88142999
A;Accession: A54498
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1701 <PET>
A;Cross-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match             57.8%;  Score 1145;  DB 2;  Length 1701;
  Best Local Similarity   56.8%;  Pred. No. 5.8e-53;
  Matches  222;  Conservative   61;  Mismatches   86;  Indels   22;  Gaps    3;

Qy       7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
           :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db    1313 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 1372

Qy      67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db    1373 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 1432

Qy     127 SEKYKSDLDSIKKYI--------------------NDKQGENEKYLPFLNNIETLYKTVN 166
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db    1433 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 1492

Qy     167 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 226
           :||| ::|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db    1493 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 1552

Qy     227 NHNNLLTKFLSTGMV--FENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRH 284
             ::| | ||||:|  | |   :::| |:: |  : : : :     |    ::||||||
Db    1553 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRH 1608

Qy     285 LDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKP 344
           |||||||||||||||||| |||||||||||||||| ||||||||: ||||||||||
Db    1609 LDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKP 1668

Db    1313 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 1372

Qy      67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db    1373 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 1432

Qy     127 SEKYKSDLDSIKKYI--------------------NDKQGENEKYLPFLNNIETLYKTVN 166
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db    1433 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 1492

Qy     167 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 226
           :||| ::|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db    1493 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 1552

Qy     227 NHNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLD 286
             ::| | ||||:| :|   :::| |:: |  : : : :     |    ::||||||||
Db    1553 TKKDMLGKLLSTGLVQIFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLD 1610

Qy     287 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDC 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1611 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDS 1670

Qy     347 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 377
           |||||||||||||||||||||||||||||||
Db    1671 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 1701


RESULT   9
A26868
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
C;Accession: A26868
R;Tanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.
J. Mol. Biol. 195, 273-287, 1987
A;Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmod
A;Reference number: A26868; MUID:88011243
A;Accession: A26868
A;Molecule type: DNA
A;Residues: 1-1701 <TAN>
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1701/Product: major merozoite surface antigen #status predicted <MAT>


  Query Match             57.1%;  Score 1131;  DB 2;  Length 1701;
  Best Local Similarity   56.5%;  Pred. No. 3.2e-52;
  Matches  222;  Conservative   60;  Mismatches   85;  Indels   26;  Gaps    4;

Qy       7 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 66
           :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db    1313 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 1372

Qy      67 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 126
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db    1373 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 1432

Qy     127 SEKYKSDLDSIKKYI--------------------NDKQGENEKYLPFLNNIETLYKTVN 166
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db    1433 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 1492

Qy     167 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 226
           :||| ::|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db    1493 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 1552

Qy     227 NHNNLLTKFLSTGMV--FENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRH 284
             ::| | ||||:|  | |   :::| |:: |  : : : :     |    ::||||||
Db    1553 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRH 1608

Qy     285 LDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKP 344
           |||||||||||||||||| |||||||||||||||| ||||||||: ||||||||||
Db    1609 LDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKP 1668
```

```
Qy    345 DCYPLFDGIFCSSSNFLGISFLLILMLILYSFI 377
          | ||||||||||||||||||||||||||||||||
Db   1669 DSYPLFDGIFCSSSNFLGISFLLLILMLILYSFI 1701
```

RESULT 10
A39401
merozoite surface antigen 1 precursor - Plasmodium vivax
C;Species: Plasmodium vivax
C;Date: 28-Feb-1992 #sequence_revision 28-Feb-1992 #text_change 24-Nov-1999
C;Accession: A39401
R;del Portillo, H.A.; Longacre, S.; Khouri, E.; David, P.H.
Proc. Natl. Acad. Sci. U.S.A. 88, 4030-4034, 1991
A;Title: Primary structure of the merozoite surface antigen 1 of Plasmodium vivax reveal
A;Reference number: A39401; MUID:91219506
A;Accession: A39401
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1726 <DEL>
A;Cross-references: GB:M60807
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

```
  Query Match             37.6%;  Score 745;  DB 2;  Length 1726;
  Best Local Similarity   39.4%;  Pred. No. 7.3e-32;
  Matches  154;  Conservative   72;  Mismatches  117;  Indels   48;  Gaps    6;

Qy      16 NEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVLESDLIPYKD 75
           ::|:|:|||||||:|: :||||||:|  || |: |:|:||  ||  |  || ||| |:|
Db    1346 SDYDVVYLKPLAGMYKKIKKQLENHVNAFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKY 1405

Qy      76 LTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSEKYKSDL- 134
             |  |::||||| |: ||: | | || ||  ||| |:  ||| : ||  : | ||: |
Db    1406 SPSGEYIIKDPYKLLDLEKKKKLLGSYKYIGASIDKDLGTANDGVNYYNKMGELYKTHLT 1465

Qy     135 -----------------DSIKKYIND------------KQGENEKYLPFLNNIETLYKTV 165
                            | |||  :|            |: | ||||||||:::  |:::
Db    1466 AVNEEVKKVEADIKAEDDKIKKIGSDSTKTTEKTQSMAKKAELEKYLPFLNSLQKEYESL 1525

Qy     166 NDKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTD 225
           |:: :  :|:  : |   ||   |: :|:|     : :|| ::||       ::   :
Db    1526 VSKVNTYTDNLKKVINNCQLEKKEAEITVKKLQDYNKMDEKLEEYKK-------SEKKNE 1578

Qy     226 YNHNNLLTKFLSTGMVFENLLKSVLSNLLDWK---LARYVKHFTTPMRKKTMIQQSSGCF 282
           : || | : : :: ||  | :|| ||: :   |     :|        | :  :: |:
Db    1579 VKSSGLLEKLMKSKLIKENESKEILSQLLNVQTQLLTMSSEHTCI----DTNVPDNAACY 1634

Qy     283 RHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECT 342
           |:||  || :||| :|:|| |||   | || :|||||  :|:|   ||     || |:||
Db    1635 RYLDGMEEWRCLLTFKEEGGKCVPGSNVTCKDNNGGCAPEAECKMTDS----NKIVCKCT 1690

Qy     343 KPDCYPLFDGIFCSSSNFLGISFLLILMLIL 373
           |    |||:|:|||||:|| :||||:::| |
Db    1691 KEGSEPLFEGVFCSSSSFLSLSFLLLMLLFL 1721
```

RESULT 11
A45604
major blood-stage surface antigen Pv200 - Plasmodium vivax
C;Species: Plasmodium vivax
C;Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
C;Accession: A45604
R;Gibson, H.L.; Tucker, J.E.; Kaslow, D.C.; Krettli, A.U.; Collins, W.E.; Kiefer, M.C.;
Mol. Biochem. Parasitol. 50, 325-333, 1992
A;Title: Structure and expression of the gene for Pv200, a major blood-stage surface ant
A;Reference number: A45604; MUID:92158013
A;Accession: A45604
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1751 <GIB>
A;Cross-references: GB:M75674; NID:g160608; PID:g457336
A;Note: sequence extracted from NCBI backbone (NCBIN:83591, NCBIP:83592)
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

```
  Query Match             37.5%;  Score 744;  DB 2;  Length 1751;
  Best Local Similarity   38.9%;  Pred. No. 8.4e-32;
  Matches  152;  Conservative   76;  Mismatches  115;  Indels   48;  Gaps    6;

Qy      16 NEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVLESDLIPYKD 75
           ::|:|:|||||||:|:::|||||:|  || |: |:|:||  ||  |  || ||| |:|
Db    1371 SDYDVVYLKPLAGMYKTIKKQLENHVNAFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKY 1430

Qy      76 LTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSEKYKSDLD 135
             :|  |::||||| |: ||: | : || ||  ||| |:  ||| : ||  : | ||: ||
Db    1431 SSSGEYIIKDPYKLLDLEKKKKLIGSYKYIGASIDMDLATANDGVTYYNKMGELYKTHLD 1490

Qy     136 SIK---KYIND--------------------------KQGENEKYLPFLNNIETLYKTV 165
             :|   | : |                          |: | ||||||||:::  |:::
Db    1491 GVKTEIKKVEDDIKKQDEELKKLGNVNSQDSKKNEFIAKKAELEKYLPFLNSLQKEYESL 1550

Qy     166 NDKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTD 225
             |:: :  :|:  : |   ||   |: :|:|     : :|| ::||       ::   :
Db    1551 VSKVNTYTDNLKKVINNCQLEKKEAEITVKKLQDYNKMDEKLEEYKK-------SEKKNE 1603

Qy     226 YNHNNLLTKFLSTGMVFENLLKSVLSNLLDWK---LARYVKHFTTPMRKKTMIQQSSGCF 282
           : || | : : :: ||  | :|| ||: :   |     :|        | :  :: |:
Db    1604 VKSSGLLEKLMKSKLIKENESKEILSQLLNVQTQLLTMSSEHTCI----DTNVPDNAACY 1659

Qy     283 RHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECT 342
           |:||  || :||| :|:|| |||   | || :|||||  :|:|   ||     || |:||
Db    1660 RYLDGTEEWRCLLTFKEEGGKCVPASNVTCKDNNGGCAPEAECKMTDS----NKIVCKCT 1715

Qy     343 KPDCYPLFDGIFCSSSNFLGISFLLILMLIL 373
           |    |||:|:|||||:|| :||||:::| |
Db    1716 KEGSEPLFEGVFCSSSSFLSLSFLLLMLLFL 1746
```

RESULT 12
A28121
major merozoite surface antigen - Plasmodium yoelii (fragment)
C;Species: Plasmodium yoelii
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 26-Aug-1999
C;Accession: A28121
R;Burns Jr., J.M.; Daly, T.M.; Vaidya, A.B.; Long, C.A.
Proc. Natl. Acad. Sci. U.S.A. 85, 602-606, 1988
A;Title: The 3' portion of the gene for a Plasmodium yoelii merozoite surface antigen
A;Reference number: A28121; MUID:88124889
A;Accession: A28121
A;Molecule type: DNA
A;Residues: 1-680 <BUR>
A;Cross-references: GB:J03612; NID:g160678; PID:g160679
A;Experimental source: strain 17XL
A;Note: the authors translated the codon GTA for residue 429 as Leu
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

```
  Query Match             29.7%;  Score 588.5;  DB 2;  Length 680;
  Best Local Similarity   34.4%;  Pred. No. 4.6e-24;
  Matches  132;  Conservative   79;  Mismatches  142;  Indels   31;  Gaps    7;

Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVLES 68
           :|||:  ||   :| | |   |:|||| :     |   :: : ||::  || :|  ||
Db     309 DILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQKRNDFLEVLSH 368

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
           :|  :|||::: ||:::||: |: :|:|| : :  |    |: ||    | : ::  : |
Db     369 ELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFNKMVE 428

Qy     129 KYKSDLDSIKKYI-------ND-KQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           | : | ::|: |       ||  : | :||:| | ::: ||:|| : : :   |: ::
Db     429 VYNTQLAAVKEQIATIEAETNDTNKEEKKKYIPILEDLKGLYETVIGQAEEYSEELQNRL 488
```

```
Qy    181 LNYTYEKSNVEVKIKELNYLKTIQDKLADF----KKNNNFVGIADLSTDYNHNNLLTKFL 236
          ||  ||:  |:  | |     | :|| :|    : | :   ||           |
Db    489 DNYKNEKAEFEILTKNLEKYIQIDEKLDEFVEHAENNKHIASIA----------LNNLN 537

Qy    237 STGMVFENLLKSVLSNLLDWK----LARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECK 292
          :|:| |   | :|: :|:      |    ||     |    | :::|||| :  || :
Db    538 KSGLVGEGESKKILAKMLNMDGMDLLGVDPKHVCVDTRD---IPKNAGCFRDDNGTEEWR 594

Qy    293 CLLNYKQ-EGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFD 351
          ||| ||: ||: |||| |||: |||||  | |   :|  | ||| | | :|     ::
Db    595 CLLGYKKGEGNTCVENNNPTCDINNGGCDPTASCQNAESTENSKKIICTCKEPTPNAYYE 654

Qy    352 GIFCSSSNFLGISFLLILMLILYS 375
          |:|||||:|:|:| |||: ||:::
Db    655 GVFCSSSSFMGLSILLIITLIVFN 678


RESULT  13
A45532
major merozoite surface antigen precursor - Plasmodium yoelii
C;Species: Plasmodium yoelii
C;Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 26-Aug-1999
C;Accession: A45532; A45531
R;Lewis, A.P.
Mol. Biochem. Parasitol. 36, 271-282, 1989
A;Title: Cloning and analysis of the gene encoding the 230-kilodalton merozoite surface
A;Reference number: A45532; MUID:90014981
A;Accession: A45532
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1772 <LEW>
A;Cross-references: GB:J04668; NID:g160492; PID:g160493
R;Daly, T.M.; Burns Jr., J.M.; Long, C.A.
Mol. Biochem. Parasitol. 36, 283-285, 1989
A;Title: Precursor to the major merozoite surface antigen of Plasmodium yoelii: cloning
A;Reference number: A45531; MUID:90014982
A;Accession: A45531
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 454-1094 <DAL>
A;Cross-references: GB:J03975; NID:g160081; PID:g160082
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match             29.7%;  Score 588.5;  DB 2;  Length 1772;
  Best Local Similarity   34.4%;  Pred. No. 1.4e-23;
  Matches  132;  Conservative   79;  Mismatches  142;  Indels   31;  Gaps    7;

Qy      9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVLES 68
          :|||:  ||   :| | |   |:|||| :     |   :: : ||::  || :|  ||
Db   1401 DILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQKRNDFLEVLSH 1460

Qy     69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
          :|  :|||::: ||:::||: |: :|:|| : :  |    |: ||    | : ::  : |
Db   1461 ELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFNKMVE 1520

Qy    129 KYKSDLDSIKKYI-------ND-KQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          | : | ::|: |       ||  : | :||:| | ::: ||:||  : : :   |: ::
Db   1521 LYNTQLAAVKEQIATIEAETNDTNKEEKKKYIPILEDLKGLYETVIGQAEEYSEELQNRL 1580

Qy    181 LNYTYEKSNVEVKIKELNYLKTIQDKLADF----KKNNNFVGIADLSTDYNHNNLLTKFL 236
          ||  ||:  |:  | |     | :|| :|    : | :   ||           |
Db   1581 DNYKNEKAEFEILTKNLEKYIQIDEKLDEFVEHAENNKHIASIA----------LNNLN 1629

Qy    237 STGMVFENLLKSVLSNLLDWK----LARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECK 292
          :|:| |   | :|: :|:      |    ||     |    | :::|||| :  || :
Db   1630 KSGLVGEGESKKILAKMLNMDGMDLLGVDPKHVCVDTRD---IPKNAGCFRDDNGTEEWR 1686

Qy    293 CLLNYKQ-EGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFD 351
          ||| ||: ||: |||| |||: |||||  | |   :|  | ||| | | :|     ::
Db   1687 CLLGYKKGEGNTCVENNNPTCDINNGGCDPTASCQNAESTENSKKIICTCKEPTPNAYYE 1746

Qy    352 GIFCSSSNFLGISFLLILMLILYS 375
          |:|||||:|:|:| |||: ||:::
Db   1747 GVFCSSSSFMGLSILLIITLIVFN 1770


RESULT  14
A45546
major merozoite surface antigen precursor - Plasmodium chabaudi chabaudi
C;Species: Plasmodium chabaudi chabaudi
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
C;Accession: A45546
R;Deleersnijder, W.; Hendrix, D.; Bendahman, N.; Hanegreefs, J.; Brijs, L.; Hamers-Ca
Mol. Biochem. Parasitol. 43, 231-244, 1990
A;Title: Molecular cloning and sequence analysis of the gene encoding the major meroz
A;Reference number: A45546; MUID:91218805
A;Accession: A45546
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1785 <DEL>
A;Cross-references: GB:M34947; NID:g160597; PID:g160598
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match             27.5%;  Score 546;  DB 2;  Length 1785;
  Best Local Similarity   30.8%;  Pred. No. 2.4e-21;
  Matches  119;  Conservative   83;  Mismatches  160;  Indels   24;  Gaps    5;

Qy      3 TTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENF 62
          |   : :||   ::| | :| | |   |:| || :          ::   || :  || :|
Db   1409 TDEEVKDILDAFKSENEYIYTKSLGNTYKSFKKHMLKEFSMIKEDIMTGLNYKLEKRNDF 1468

Qy     63 KNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGY 122
           :||  :|  :||: :: :|||:||: |: :|:|| : :  |    :  ||  | | : :
Db   1469 LDVLSYELALFKDINTNKFVVKNPYQLLDNDKKDKQMINLKYAIKGVTEDIETATDGIEF 1528

Qy    123 YKILSEKYKSDLDSIKKYI-----NDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLE 177
          :  : | ||  |::: : |          | :|| |   ::: ||:|: :  : |   |:
Db   1529 FNKMIELYKPQLNAVNEQIAAIGTEPTDAEKKKYAPIFEDLKGLYETILNGAEEFSELLQ 1588

Qy    178 AKVLNYTYEKSNVEVKIKELNYLKTIQDKLADF----KKNNNFVGIADLSTDYNHNNLLT 233
           |: ||  ||:  :: :  |     | :|| ||    :|| :   ||           |
Db   1589 HKLENYKIEKAGFDILMANLETYIRIDEKLEDFVESAEKNKHIASIA----------LN 1637

Qy    234 KFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKC 293
          :|:| |   | :|: :|:      :   :  :   |    ::|||: |  || :|
Db   1638 NLNKSGLVTEGESKKILAKMLNMDAMDLLGIGSNHVCISTSTPDNAGCFRYDDGTEEWRC 1697

Qy    294 LLNYKQ--EGDKCVENPNPTCNENNGGCDADAKCTEEDSGSN--GKKITCECTKPDCYPL 349
          || :|:  :|::|| :  | || |||||| :| | | ::      ||| | | :|:
Db   1698 LLGFKKDDDGNRCVADDAPVCNNNNGGCDKNADCREVENTDRDPSKKIVCTCKEPNPNAY 1757

Qy    350 FDGIFCSSSNFLGISFLLILMLILYS 375
          : |:||||| |:|:| |||: ||:::
Db   1758 YAGVFCSSSGFMGLSILLIITLIVFN 1783


RESULT  15
T18501
hypothetical protein C0760c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18501
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18501
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3394 <LAW>
A;Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331922; PIDN:CAB11140.1
C;Genetics:
```

A;Map position: 3
A;Note: C0760c

Query Match 7.9%; Score 157.5; DB 2; Length 3394;
Best Local Similarity 24.8%; Pred. No. 1.5;
Matches 90; Conservative 61; Mismatches 121; Indels 91; Gaps 22;

```
Qy      8 DNILSKIENEYEVLYLKPLAGVYRSLK------------KQLENNVMTFNVNVKDILNS 54
          | || ::  :|| |    |   | :|              |:|| |:  :| | | ||:
Db   1201 DKIL-RLTKKYEQL----LEENYEKIKNKNEKEEYLHACIKELEMNLERYN-NEKIILDE 1254

Qy     55 RFNKRENFKNVL-ESDLI---PYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSID 110
          : |::|   |:: |  ||    |::  : |  :    : | |||  | |      || |:
Db   1255 QINEKEKKINIINEKYLILEKEYEEYQNKNIFINAQIENLEKEK--KQLQEEIIQKDMIN 1312

Qy    111 TDINFAN-DVLGYYKILSEKYKSDLDSI-------KKYINDK----QGENEKYLPF---- 154
            :|  | |::  |    || |  | ::         |: ||       ||| | :
Db   1313 VKLNEKNCDIIKIY----EKEKQYLHTLLQENKDSHNYLKDKFENLLNLNEK-LKYDHDI 1367

Qy    155 -LNNIETLY--KTVNDKIDLFVIHLEAKVLNYTYEKSNVEVKIKEL-NYLKTIQDKLAD- 209
           || | ||:  :  |:| : |  |:     |   | :|: :|:|||| |    |: :| :
Db   1368 SLNKINTLWLEEKENNKKNTF--HMN----NLRVENNNLLLKMKELQNKYNIIKKELNER 1421

Qy    210 ------FKKNNNFVGIAD----------LSTDYNHNNLLTKFLSTGMVFENLLKSVLSNL 253
              |: | : : : |          ::  | :|  |  |:    :  |:     ||:
Db   1422 IKQINVFRNNVSTLSLRDNRSTRGSIHQINNMYMNNTHL-GFMGASKINNNISNLYYSNM 1480

Qy    254 LDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCN 313
          :      :: |  : :: |   : :|            :  :| |   |  : | : | |
Db   1481 I------HMSHRGSIIKNKEDAEGNS----------TQARMNNKDSTDNIINNIHNTDN 1523

Qy    314 ENN 316
            ||
Db   1524 INN 1526
```

Search completed: August 8, 2001, 12:31:02
Job time: 110 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:30:01 ; Search time 34.41 Seconds
(without alignments)
225.590 Million cell updates/sec

Title: US-09-500-376-5
Perfect score: 1982
Sequence: 1 AVTTSVIDNILSKIENEYEV.....SNFLGISFLLILMLILYSFI 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	296	14.9	53	1	US-08-290-919-3	Sequence 3, Appli
2	277	14.0	53	1	US-08-290-919-4	Sequence 4, Appli
3	243	12.3	106	1	US-08-290-919-11	Sequence 11, Appl
4	201	10.1	48	1	US-08-290-919-1	Sequence 1, Appli
5	201	10.1	48	1	US-08-290-919-12	Sequence 12, Appl
6	198	10.0	48	1	US-08-290-919-2	Sequence 2, Appli
7	125.5	6.3	2391	2	US-08-446-855A-2	Sequence 2, Appli
8	125.5	6.3	2391	4	US-09-150-741-2	Sequence 2, Appli
9	124	6.3	1786	4	US-08-973-462-8	Sequence 8, Appli
10	122	6.2	3135	1	US-08-323-170B-2	Sequence 2, Appli
11	121.5	6.1	984	1	US-08-257-073-3	Sequence 3, Appli
12	121.5	6.1	984	2	US-08-184-009-120	Sequence 120, App
13	121.5	6.1	984	2	US-08-458-356-120	Sequence 120, App
14	119.5	6.0	1435	2	US-08-568-459A-4	Sequence 4, Appli
15	119.5	6.0	1435	2	US-08-487-826B-4	Sequence 4, Appli
16	119.5	6.0	1494	3	US-08-755-587-186	Sequence 186, App
17	118	6.0	978	2	US-08-415-593-43	Sequence 43, Appl
18	117	5.9	455	2	US-08-392-625-21	Sequence 21, Appl
19	117	5.9	455	2	US-08-466-961A-21	Sequence 21, Appl
20	117	5.9	455	2	US-08-645-193B-23	Sequence 23, Appl
21	116.5	5.9	1147	3	US-08-470-260-5	Sequence 5, Appli
22	116.5	5.9	1147	3	US-08-471-491-5	Sequence 5, Appli
23	116.5	5.9	1147	4	US-08-466-662-5	Sequence 5, Appli
24	116.5	5.9	3289	2	US-08-477-451-2	Sequence 2, Appli
25	113.5	5.7	1239	2	US-08-937-931-2	Sequence 2, Appli
26	113.5	5.7	1239	4	US-09-285-502-2	Sequence 2, Appli
27	112.5	5.7	897	4	US-08-974-549A-189	Sequence 189, App
28	112.5	5.7	1007	3	US-08-851-843A-86	Sequence 86, Appl
29	112.5	5.7	1007	4	US-08-974-549A-187	Sequence 187, App
30	112.5	5.7	1031	3	US-08-851-843A-2	Sequence 2, Appli
31	112.5	5.7	1031	4	US-08-974-549A-110	Sequence 110, App
32	110.5	5.6	2710	1	US-08-480-604A-6	Sequence 6, Appli
33	110.5	5.6	2710	2	US-08-405-496A-6	Sequence 6, Appli
34	109.5	5.5	990	2	US-08-392-625-20	Sequence 20, Appl
35	109.5	5.5	990	2	US-08-466-961A-20	Sequence 20, Appl
36	109.5	5.5	990	2	US-08-645-193B-15	Sequence 15, Appl
37	109.5	5.5	2366	1	US-08-480-604A-10	Sequence 10, Appl
38	109.5	5.5	2366	2	US-08-405-496A-10	Sequence 10, Appl
39	108.5	5.5	903	2	US-08-209-521-24	Sequence 24, Appl
40	108.5	5.5	903	4	US-08-961-810-134	Sequence 134, App
41	108.5	5.5	903	4	US-08-352-902D-134	Sequence 134, App
42	108	5.4	439	2	US-08-853-659A-37	Sequence 37, Appl
43	107	5.4	778	5	PCT-US93-03076-4	Sequence 4, Appli
44	106.5	5.4	599	2	US-08-910-551B-2	Sequence 2, Appli
45	106.5	5.4	720	2	US-08-840-236-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-290-919-3
; Sequence 3, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
;   APPLICANT: HOLDER, ANTHONY A.
;   APPLICANT: BLACKMAN, MICHAEL J.
;   APPLICANT: CHAPPEL, JONATHAN A.
;   TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
;   TITLE OF INVENTION: VACCINE
;   NUMBER OF SEQUENCES: 19
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
;     STREET: 1100 NEW YORK AVENUE, N.W.
;     CITY: WASHINGTON
;     STATE: D.C.
;     COUNTRY: U.S.A.
;     ZIP: 20005-3918
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/290,919
;     FILING DATE: 04-OCT-1994
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: GB 9203821.5
;     FILING DATE: 22-FEB-1992
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: PCT/GB93/00367
;     FILING DATE: 22-FEB-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: KOKULIS, PAUL N.
;     REGISTRATION NUMBER: 16,773
;     REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (202) 861-3000
;     TELEFAX: (202) 822-0944
;     TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 53 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   FEATURE:
;     NAME/KEY: Modified-site
```

```
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-3


  Query Match             14.9%;  Score 296;  DB 1;  Length 53;
  Best Local Similarity   98.1%;  Pred. No. 1.8e-16;
  Matches   51;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     308 PNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSN 359
           |||||||||||||||||||||||||||||||||||| ||||||||||||
Db       2 PNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSN 53


RESULT   2
US-08-290-919-4
; Sequence 4, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-4


  Query Match             14.0%;  Score 277;  DB 1;  Length 53;
  Best Local Similarity   92.3%;  Pred. No. 5.3e-15;
  Matches   48;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy     308 PNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSN 359
           |||||||||||||||| ||||||||: |||||||||| ||||||||||||
Db       2 PNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDSYPLFDGIFCSSSN 53


RESULT   3
US-08-290-919-11
; Sequence 11, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  106 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-290-919-11


  Query Match             12.3%;  Score 243;  DB 1;  Length 106;
  Best Local Similarity   48.0%;  Pred. No. 5.4e-12;
  Matches   47;  Conservative   14;  Mismatches   33;  Indels    4;  Gaps    2;

Qy     263 KHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQ-EGDKCVENPNPTCNENNGGCDA 321
           ||      |    | :::|||| :  || :||| ||: ||: |||| ||||: ||||||
Db      12 KHVCVDTRD---IPKNAGCFRDDNGTEEWRCLLGYKKGEGNTCVENNNPTCDINNGGCDP 68

Qy     322 DAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSN 359
           | |   :|  | ||| | | :|     ::|:|||||:
Db      69 TASCQNAESTENSKKIICTCKEPTPNAYYEGVFCSSSS 106


RESULT   4
US-08-290-919-1
```

```
; Sequence 1, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
;   APPLICANT:  HOLDER, ANTHONY A.
;   APPLICANT:  BLACKMAN, MICHAEL J.
;   APPLICANT:  CHAPPEL, JONATHAN A.
;   TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;   TITLE OF INVENTION:  VACCINE
;   NUMBER OF SEQUENCES:  19
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;     STREET:  1100 NEW YORK AVENUE, N.W.
;     CITY:  WASHINGTON
;     STATE:  D.C.
;     COUNTRY:  U.S.A.
;     ZIP:  20005-3918
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/290,919
;     FILING DATE:  04-OCT-1994
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  GB 9203821.5
;     FILING DATE:  22-FEB-1992
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  PCT/GB93/00367
;     FILING DATE:  22-FEB-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  KOKULIS, PAUL N.
;     REGISTRATION NUMBER:  16,773
;     REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (202) 861-3000
;     TELEFAX:  (202) 822-0944
;     TELEX:  6714627 CUSH
; INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  48 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
;   FEATURE:
;     NAME/KEY:  Modified-site
;     LOCATION:  1
;     OTHER INFORMATION:  /label= X
;     OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-1


  Query Match             10.1%;  Score 201;  DB 1;  Length 48;
  Best Local Similarity   87.5%;  Pred. No. 3.3e-09;
  Matches   35;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    272 KTMIQQSSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 311
          |    |:||||||||||||||||||||||||||||||||||
Db      9 KKQCPQNSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 48


RESULT   5
US-08-290-919-12
; Sequence 12, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
;   APPLICANT:  HOLDER, ANTHONY A.
;   APPLICANT:  BLACKMAN, MICHAEL J.
;   APPLICANT:  CHAPPEL, JONATHAN A.
;   TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;   TITLE OF INVENTION:  VACCINE
;   NUMBER OF SEQUENCES:  19
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;     STREET:  1100 NEW YORK AVENUE, N.W.
;     CITY:  WASHINGTON
;     STATE:  D.C.
;     COUNTRY:  U.S.A.
;     ZIP:  20005-3918
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/290,919
;     FILING DATE:  04-OCT-1994
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  GB 9203821.5
;     FILING DATE:  22-FEB-1992
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  PCT/GB93/00367
;     FILING DATE:  22-FEB-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  KOKULIS, PAUL N.
;     REGISTRATION NUMBER:  16,773
;     REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (202) 861-3000
;     TELEFAX:  (202) 822-0944
;     TELEX:  6714627 CUSH
; INFORMATION FOR SEQ ID NO:  12:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  48 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  peptide
US-08-290-919-12


  Query Match             10.1%;  Score 201;  DB 1;  Length 48;
  Best Local Similarity   87.5%;  Pred. No. 3.3e-09;
  Matches   35;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    272 KTMIQQSSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 311
          |    |:||||||||||||||||||||||||||||||||||
Db      9 KKQCPQNSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 48


RESULT   6
US-08-290-919-2
; Sequence 2, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
;   APPLICANT:  HOLDER, ANTHONY A.
;   APPLICANT:  BLACKMAN, MICHAEL J.
;   APPLICANT:  CHAPPEL, JONATHAN A.
;   TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;   TITLE OF INVENTION:  VACCINE
;   NUMBER OF SEQUENCES:  19
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;     STREET:  1100 NEW YORK AVENUE, N.W.
;     CITY:  WASHINGTON
;     STATE:  D.C.
;     COUNTRY:  U.S.A.
;     ZIP:  20005-3918
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  48 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-2


  Query Match             10.0%;  Score 198;  DB 1;  Length 48;
  Best Local Similarity   85.0%;  Pred. No. 5.6e-09;
  Matches   34;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy     272 KTMIQQSSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 311
           |    ::|||||||||||||||||||||||||||||||||||
Db       9 KKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 48


RESULT   7
US-08-446-855A-2
; Sequence 2, Application US/08446855A
; Patent No. 5849573
;  GENERAL INFORMATION:
;    APPLICANT:  Stewart, Thomas S
;    APPLICANT:  Flores, Maria V
;    APPLICANT:  O'Sullivan, William J
;    TITLE OF INVENTION:  Nucleotide sequence encoding carbamoyl
;    TITLE OF INVENTION:  phosphate synthetase II
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye PC
;      STREET:  1100 No. 5849573th Glebe Road, 8th Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,855A
;      FILING DATE:  06-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mitchard, Leonard C
;      REGISTRATION NUMBER:  29.009
;      REFERENCE/DOCKET NUMBER:  47-80
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4000
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2391 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-446-855A-2


  Query Match             6.3%;  Score 125.5;  DB 2;  Length 2391;
  Best Local Similarity   19.3%;  Pred. No. 0.37;
  Matches   68;  Conservative   62;  Mismatches  111;  Indels  111;  Gaps   15;

Qy      37 LENNV---MTFNVNVKDILNSRFNKRENFKNVLESDLIPYKDLTSSNYVV--------KD 85
           :|||:   :: | |:|  :|:  |   | :|| |::|    :   |:::         |
Db    1333 IENNICHDISINKNIKVTINNSNNSISNNENV-ETNLNCVSERAGSHHIYGKEEKSIGSD 1391

Qy      86 PYKFLNKEKRDKFLSSYNYIKDSIDTDIN-FANDVLGYYKILSEKYKSDLDSIKKY---- 140
             |: :  :  |  |   :  : |:|   ||          | : |:::   :
Db    1392 DTNILSAQNSNNNFSCNNENMNKANVDVNVLENDT---------KKREDINTTTVFMEGQ 1442

Qy     141 ---INDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKVLNYTYEKSNVEVKIKEL 197
             ||:|  ||   |              |: |: :::|: |  ||    :||: : |::
Db    1443 NSVINNKNKENSSLLK----------GDEEDIVMVNLK-KENNYNSVINNVDCRKKDM 1489

Qy     198 NYLKTIQDKLADFKK--------NNNFVGIADLSTDYNHNNLLTKFLSTGMVFENLLKSV 249
           :  | | |:   :||        |||  |      | :: |: |
Db    1490 DG-KNINDECKTYKKNKYKDMGLNNNIVDELSNGTSHSTNDHL---------------- 1531

Qy     250 LSNLLDWKLARYVKHFTTP-------------MRKKTMIQQSSG--------CFRHLDE 287
                     |: :| |                ::|::  :: |         : :  :
Db    1532 ----------YLDNFNTSDEEIGNNKNMDMYLSKEKSISNKNPGNSYYVVDSVYNNEYK 1580

Qy     288 REECKCLLNYKQEGDKCVENPNPTC-NENNGGCDADAK---------CTEED 329
             : | |:: :   |:   | |  | | ||     : |         | |::
Db    1581 INKMKELIDNENLNDEYNNNVNMNCSNYNNASAFVNGKDRNDNLENDCIEKN 1632


RESULT   8
US-09-150-741-2
; Sequence 2, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
;  APPLICANT: Stewart et al.
;  TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
;  TITLE OF INVENTION: Synthetase II
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/150,741
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PL6380
;  EARLIER FILING DATE: 1992-12-16
;  EARLIER APPLICATION NUMBER: AU93/00617
;  EARLIER FILING DATE: 1993-12-02
;  EARLIER APPLICATION NUMBER: 08/446,855
;  EARLIER FILING DATE: 1995-07-06
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2391
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2


  Query Match             6.3%;  Score 125.5;  DB 4;  Length 2391;
  Best Local Similarity   19.3%;  Pred. No. 0.37;
```

```
  Matches   68;  Conservative   62;  Mismatches  111;  Indels  111;  Gaps   15;

Qy      37 LENNV---MTFNVNVKDILNSRFNKRENFKNVLESDLIPYKDLTSSNYVV--------KD 85
           :|||:   :: | |:|  :|:  |   | :|| |::|    :   |:::           |
Db    1333 IENNICHDISINKNIKVTINNSNNSISNNENV-ETNLNCVSERAGSHHIYGKEEKSIGSD 1391

Qy      86 PYKFLNKEKRDKFLSSYNYIKDSIDTDIN-FANDVLGYYKILSEKYKSDLDSIKKY---- 140
             |: :  :   |  |   :  : |:|   ||          | : |:::   :
Db    1392 DTNILSAQNSNNNFSCNNENMNKANVDVNVLENDT---------KKREDINTTTVFMEGQ 1442

Qy     141 ---INDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKVLNYTYEKSNVEVKIKEL 197
              ||:|  ||   |              |: |: :::|: |  ||    :||: : |:|
Db    1443 NSVINNKNKENSSLLK-----------GDEEDIVMVNLK-KENNYNSVINNVDCRKKDM 1489

Qy     198 NYLKTIQDKLADFKK--------NNNFVGIADLSTDYNHNNLLTKFLSTGMVFENLLKSV 249
           :  | | |:   :||        ||| |      | :: |: |
Db    1490 DG-KNINDECKTYKKNKYKDMGLNNNIVDELSNGTSHSTNDHL---------------- 1531

Qy     250 LSNLLDWKLARYVKHFTTP--------------MRKKTMIQQSSG--------CFRHLDE 287
                    |: :| |                 ::|::  :: |          : :  :
Db    1532 -----------YLDNFNTSDEEIGNNKNMDMYLSKEKSISNKNPGNSYYVVDSVYNNEYK 1580

Qy     288 REECKCLLNYKQEGDKCVENPNPTC-NENNGGCDADAK---------CTEED 329
             : | |:: :   |:   | |  | | ||     : |          | |::
Db    1581 INKMKELIDNENLNDEYNNNVNMNCSNYNNASAFVNGKDRNDNLENDCIEKN 1632


RESULT   9
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT: DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 0660-0125-0 PCT
;  CURRENT APPLICATION NUMBER: US/08/973,462B
;  CURRENT FILING DATE: 1998-02-06
;  EARLIER APPLICATION NUMBER: PCT/FR96/00894
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: FR 95/07007
;  EARLIER FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 1786
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Polypeptide
US-08-973-462-8


  Query Match             6.3%;  Score 124;  DB 4;  Length 1786;
  Best Local Similarity   25.3%;  Pred. No. 0.33;
  Matches   68;  Conservative   45;  Mismatches   80;  Indels   76;  Gaps   17;

Qy       2 VTTSVIDNI-------LSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNV---KD- 50
           | :|::|||       |:|:||      :    ||  :: : :| ||   :|:|   ||
Db    1047 VVSSILDNIENMKEGLLNKLEN------ISSTEGVQETVTEHVEQNVYV-DVDVPAMKDQ 1099

Qy      51 ---ILNSR-------FNKRENFKNVLESDLIPYKDLTSSNYVVKDPYK------------ 88
              |||          ||  : ||:  |||:|  :::      | |:  |
Db    1100 FLGILNEAGGLKEMFFNLEDVFKS--ESDVITVEEI-KDEPVQKEVEKETVSIIEEMEEN 1156

Qy      89 ---FLNKEKR---DKFLSSYNYIKDSIDTDINFANDVLGYYKILSEKYKSDLDSIKKYI- 141
              | :||    || :       |:::  |  ::|     : : :| | |:  : : :
Db    1157 IVDVLEEEKEDLTDKMI-------DAVEESIEISSDSKEETESIKDKEK-DVSLVVEEVQ 1208

Qy     142 -NDKQGENEKYLPFLNNIETLYK---TVNDKIDLFVIHLEAKVLNYTYEKSNVEVK-IKE 196
            ||    || |   |  | | |    :||       : :|::  | | : ||   ||:
Db    1209 DNDMDESVEKVLELKNMEEELMKDAVEIND--------ITSKLIEETQELNEVEADLIKD 1260

Qy     197 LNYLKTIQDKLADFKKNNNFVGIADLSTD 225
           :  || ::  |::  |      | |   |
Db    1261 MEKLKELEKALSEDSKE-----IIDAKDD 1284


RESULT  10
US-08-323-170B-2
; Sequence 2, Application US/08323170B
; Patent No. 5733772
;  GENERAL INFORMATION:
;    APPLICANT:  Williamson, Kim C.
;    APPLICANT:  Kaslow, David C.
;    TITLE OF INVENTION:  Cloning and Expression of Plasmodium
;    TITLE OF INVENTION:  faliciparum Transmission-Blocking Target Antigen, Pfs230
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/323,170B
;      FILING DATE:  13-OCT-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/010,409
;      FILING DATE:  29-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Quine, Jonathan A.
;      REGISTRATION NUMBER:  P-41,261
;      REFERENCE/DOCKET NUMBER:  015280-113100US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3135 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-323-170B-2


  Query Match             6.2%;  Score 122;  DB 1;  Length 3135;
  Best Local Similarity   22.4%;  Pred. No. 0.98;
  Matches   65;  Conservative   41;  Mismatches   82;  Indels  102;  Gaps   13;

Qy      21 LYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVLESDLIPYKDLTSSN 80
           |:|  |  :: | || |  ||:  | |:||  : ||:             |  |     |:
Db      10 LFLIILVYIF-SEKKDLRCNVIKGN-NIKDDEDKRFH-----------LFYY-----SH 50

Qy      81 YVVKDPYKFLNKEKRDKFLSS---YNYIKD---SIDTDINF-----------------AN 117
             : | |     | |:: |  :   ||  |:     || :                   :
Db      51 NLFKTPETKEKKNKKECFYKNGGIYNLSKEIRMRKDTSVKIKQRTCPFHKEGSSFEMGSK 110

Qy     118 DVLGYYKILSEKYKSDLDSI---KKYIND-------KQGENEKYLPFLNNIETLYKTVND 167
           ::  :| |: :| :  ||:|   |   ||            || :  : ||:   |  ||
Db     111 NITCFYPIVGKKERKTLDTIIIKKNVTNDHVVSSDMHSNVQEKNMILIRNIDKENK--ND 168

Qy     168 KIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYN 227
                 : ::| |:   |||                                     : ||
Db     169 -----IQNVEEKIQRDTYE-----------------------------------NKDYE 187

Qy     228 HNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQ 277
             :: | ::       || | : |   |      :| |::| ||||::|:
Db     188 SDDTLIEWFDDNTNEENFLLTFLKRCL-------MKIFSSPKRKKTVVQK 230
```

```
RESULT 11
US-08-257-073-3
; Sequence 3, Application US/08257073
; Patent No. 5766597
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  de Taisne, Charles
;    APPLICANT:  Tine, John A.
;    TITLE OF INVENTION:  MALARIA RECOMBINANT POXVIRUS VACCINE
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue, 25th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/257,073
;      FILING DATE:  09-JUN-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/075,783
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/852,305
;      FILING DATE:  18-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,183
;      FILING DATE:  20-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2570
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066 CURTMS
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  984 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-257-073-3


  Query Match           6.1%;  Score 121.5;  DB 1;  Length 984;
  Best Local Similarity   22.3%;  Pred. No. 0.23;
  Matches   88;  Conservative   54;  Mismatches  149;  Indels  103;  Gaps   20;

Qy       7 IDNILSKIENEYEVLYLKPLA-----GVYRSLKKQLENNVMTFNVNVKDILNSRFNKREN 61
           : ||    :|   |:|:|          ||   |   |||     |:|:  |    |::|
Db     239 LQNICETGKNFKLVVYIKENTLIIKWKVYGETKDTTENN----KVDVRKYL---INEKET 291

Qy      62 -FKNVLESDLIPYKDLTSSNYVVKDPYKFLN--KEKRDKFLSSYNYIKDSIDTDINF--- 115
            | ::|   :  ||:   :| :    |   :   || |  |:|  ::  : : :  |
Db     292 PFTSIL---IHAYKEHNGTNLIESKNYALGSDIPEKCDT-LASNCFLSGNFNIEKCFQCA 347

Qy     116 -------ANDVLGYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIET----LYKT 164
                  |||   || |||    |    ||    |   ::         :|:     ::||
Db     348 LLVEKENKNDVC--YKYLSEDIVSKFKEIKAETEDDDEDDYTEYKLTESIDNILVKMFKT 405

Qy     165 --VNDKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQD-----KLADFKKN---- 213
             |||  :|  :           |: :  :|:: :||   ::|      | ::
Db     406 NENNDKSELIKL----------EEVDDSLKLELMNYCSLLKDVDTTGTLDNYGMGNEMD 454

Qy     214 --NNFVGIADLSTDYNHNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRK 271
             ||   :    :: | | |  || :  :         | |: ||           : |
Db     455 IFNNLKRLLIYHSEENINTLKNKFRNAAV--------CLKNVDDW----------IVNK 495

Qy     272 KTMIQQSSGCFRHLDEREECKCLLNYKQE-------GDKCVENPNPTCNENNGGCDADAK 324
           : ::          | |       |||  |       ||   :|    |:  |    |
Db     496 RGLV---------LPE-------LNYDLEYFNEHLYNDK--NSPEDKDNKGKGVVHVDTT 537

Qy     325 CTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSS 358
             :||: |        | |  |  | |   | |:
Db     538 LEKEDTLSYDNSDNMFCNKEYCNRLKDENNCISN 571


RESULT 12
US-08-184-009-120
; Sequence 120, Application US/08184009
; Patent No. 5833975
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;    NUMBER OF SEQUENCES:  217
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/184,009
;      FILING DATE:  19-JAN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2530
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  120:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  984 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-184-009-120


  Query Match           6.1%;  Score 121.5;  DB 2;  Length 984;
  Best Local Similarity   22.3%;  Pred. No. 0.23;
  Matches   88;  Conservative   54;  Mismatches  149;  Indels  103;  Gaps   20;

Qy       7 IDNILSKIENEYEVLYLKPLA-----GVYRSLKKQLENNVMTFNVNVKDILNSRFNKREN 61
           : ||    :|   |:|:|          ||   |   |||     |:|:  |    |::|
Db     239 LQNICETGKNFKLVVYIKENTLIIKWKVYGETKDTTENN----KVDVRKYL---INEKET 291

Qy      62 -FKNVLESDLIPYKDLTSSNYVVKDPYKFLN--KEKRDKFLSSYNYIKDSIDTDINF--- 115
            | ::|   :  ||:   :| :    |   :   || |  |:|  ::  : : :  |
Db     292 PFTSIL---IHAYKEHNGTNLIESKNYALGSDIPEKCDT-LASNCFLSGNFNIEKCFQCA 347
```

```
Qy    116 -------ANDVLGYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIET----LYKT 164
Db    348 LLVEKENKNDVC--YKYLSEDIVSKFKEIKAETEDDDEDDYTEYKLTESIDNILVKMFKT 405

Qy    165 --VNDKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQD-----KLADFKKN---- 213
Db    406 NENNDKSELIKL-----------EEVDDSLKLELMNYCSLLKDVDTTGTLDNYGMGNEMD 454

Qy    214 --NNFVGIADLSTDYNHNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRK 271
Db    455 IFNNLKRLLIYHSEENINTLKNKFRNAAV--------CLKNVDDW-----------IVNK 495

Qy    272 KTMIQQSSGCFRHLDEREECKCLLNYKQE-------GDKCVENPNPTCNENNGGCDADAK 324
Db    496 RGLV---------LPE-------LNYDLEYFNEHLYNDK--NSPEDKDNKGKGVVHVDTT 537

Qy    325 CTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSS 358
Db    538 LEKEDTLSYDNSDNMFCNKEYCNRLKDENNCISN 571
```

```
RESULT  13
US-08-458-356-120
; Sequence 120, Application US/08458356
; Patent No. 5942235
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;    NUMBER OF SEQUENCES:  217
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,356
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/184,009
;      FILING DATE:  19-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2530
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  120:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  984 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-458-356-120


  Query Match             6.1%;  Score 121.5;  DB 2;  Length 984;
  Best Local Similarity   22.3%;  Pred. No. 0.23;
  Matches   88;  Conservative   54;  Mismatches  149;  Indels  103;  Gaps   20;
```

```
Qy      7 IDNILSKIENEYEVLYLKPLA-----GVYRSLKKQLENNVMTFNVNVKDILNSRFNKREN 61
Db    239 LQNICETGKNFKLVVYIKENTLIIKWKVYGETKDTTENN----KVDVRKYL---INEKET 291

Qy     62 -FKNVLESDLIPYKDLTSSNYVVKDPYKFLN--KEKRDKFLSSYNYIKDSIDTDINF--- 115
Db    292 PFTSIL---IHAYKEHNGTNLIESKNYALGSDIPEKCDT-LASNCFLSGNFNIEKCFQCA 347

Qy    116 -------ANDVLGYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIET----LYKT 164
Db    348 LLVEKENKNDVC--YKYLSEDIVSKFKEIKAETEDDDEDDYTEYKLTESIDNILVKMFKT 405

Qy    165 --VNDKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQD-----KLADFKKN---- 213
Db    406 NENNDKSELIKL-----------EEVDDSLKLELMNYCSLLKDVDTTGTLDNYGMGNEMD 454

Qy    214 --NNFVGIADLSTDYNHNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRK 271
Db    455 IFNNLKRLLIYHSEENINTLKNKFRNAAV--------CLKNVDDW-----------IVNK 495

Qy    272 KTMIQQSSGCFRHLDEREECKCLLNYKQE-------GDKCVENPNPTCNENNGGCDADAK 324
Db    496 RGLV---------LPE-------LNYDLEYFNEHLYNDK--NSPEDKDNKGKGVVHVDTT 537

Qy    325 CTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSS 358
Db    538 LEKEDTLSYDNSDNMFCNKEYCNRLKDENNCISN 571
```

```
RESULT  14
US-08-568-459A-4
; Sequence 4, Application US/08568459A
; Patent No. 5849306
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/568,459A
;      FILING DATE:  07-DEC-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1435 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

```
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium falciparum
US-08-568-459A-4


  Query Match             6.0%;  Score 119.5;  DB 2;  Length 1435;
  Best Local Similarity   21.0%;  Pred. No. 0.55;
  Matches   81;  Conservative   46;  Mismatches  123;  Indels  135;  Gaps   19;

Qy      12 SKIENEYEVLYLKPLAGVYRSLKKQL----------ENNVMTFNVNVKDIL-NSRFNKR 59
           :|  |||::   :    ||:   :|          :  : ||  |  ||| ||:||||
Db      19 AKARNEYDIKENEKFLDVYKEKFNELDKKKYGNVQKTDKKIFTFIENKLDILNNSKFNKR 78

Qy      60 ----------------------ENFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRD 96
                                  | | |  :|       |::|: : :: |  :|  :
Db      79 WKSYGTPDNIDKNMSLINKHNNEEMFNNNYQS------FLSTSSLIKQNKYVPINAVRVS 132

Qy      97 KFLSSYNYIKDSIDTDINFANDVLGYYKILS---EKYKSDLDSIKKYINDKQGENEKYLP 153
           : ||       :|: ||   :     ::||   || |     ||  ||:
Db     133 RILS-------FLDSRINNGRNTSSNNEVLSNCREKRKGMKWDCKKK-NDRSN------- 177

Qy     154 FLNNIETLYKTVND-KIDLFVIHLEAKVLNYTYEKSN---VEVKIKELNYLKTIQDKLAD 209
                 | : | :| | :::|   :  || |      :|   ||   |
Db     178 --------YVCIPDRRIQLCIVNLSI-IKTYTKETMKDHFIEASKKESQLL--------- 219

Qy     210 FKKNNNFVGIADLSTDYNH---NNLLTKFLSTGMV-----------------FENLLKS 248
            |||:|          ||   |:|   ||  | :                 : : |
Db     220 LKKNDN---------KYNSKFCNDLKNSFLDYGHLAMGNDMDFGGYSTKAENKIQEVFKG 270

Qy     249 VLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE-------- 300
               : : |:  : | :    |:|      |    | :   ||   | ||
Db     271 AHGEISEHKIKNFRKEWWNEFREKLWEAMLS---EHKNNINNCK---NIPQEELQITQWI 324

Qy     301 ----GDKCVEN------PNPTCNEN 315
               |:  :|       |   |  |
Db     325 KEWHGEFLLERDNRSKLPKSKCKNN 349


RESULT  15
US-08-487-826B-4
; Sequence 4, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1435 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium falciparum
US-08-487-826B-4


  Query Match             6.0%;  Score 119.5;  DB 2;  Length 1435;
  Best Local Similarity   21.0%;  Pred. No. 0.55;
  Matches   81;  Conservative   46;  Mismatches  123;  Indels  135;  Gaps   19;

Qy      12 SKIENEYEVLYLKPLAGVYRSLKKQL----------ENNVMTFNVNVKDIL-NSRFNKR 59
           :|  |||::   :    ||:   :|          :  : ||  |  ||| ||:||||
Db      19 AKARNEYDIKENEKFLDVYKEKFNELDKKKYGNVQKTDKKIFTFIENKLDILNNSKFNKR 78

Qy      60 ----------------------ENFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRD 96
                                  | | |  :|       |::|: : :: |  :|  :
Db      79 WKSYGTPDNIDKNMSLINKHNNEEMFNNNYQS------FLSTSSLIKQNKYVPINAVRVS 132

Qy      97 KFLSSYNYIKDSIDTDINFANDVLGYYKILS---EKYKSDLDSIKKYINDKQGENEKYLP 153
           : ||       :|: ||   :     ::||   || |     ||  ||:
Db     133 RILS-------FLDSRINNGRNTSSNNEVLSNCREKRKGMKWDCKKK-NDRSN------- 177

Qy     154 FLNNIETLYKTVND-KIDLFVIHLEAKVLNYTYEKSN---VEVKIKELNYLKTIQDKLAD 209
                 | : | :| | :::|   :  || |      :|   ||   |
Db     178 --------YVCIPDRRIQLCIVNLSI-IKTYTKETMKDHFIEASKKESQLL--------- 219

Qy     210 FKKNNNFVGIADLSTDYNH---NNLLTKFLSTGMV-----------------FENLLKS 248
            |||:|          ||   |:|   ||  | :                 : : |
Db     220 LKKNDN---------KYNSKFCNDLKNSFLDYGHLAMGNDMDFGGYSTKAENKIQEVFKG 270

Qy     249 VLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE-------- 300
               : : |:  : | :    |:|      |    | :   ||   | ||
Db     271 AHGEISEHKIKNFRKEWWNEFREKLWEAMLS---EHKNNINNCK---NIPQEELQITQWI 324

Qy     301 ----GDKCVEN------PNPTCNEN 315
               |:  :|       |   |  |
Db     325 KEWHGEFLLERDNRSKLPKSKCKNN 349


Search completed: August  8, 2001, 12:30:04
Job time: 52 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:33:44 ; Search time 70.07 Seconds
(without alignments)
708.069 Million cell updates/sec

Title: US-09-500-376-4
Perfect score: 1967
Sequence: 1 AVTPSVIHNILSKIENEYEV.....SNFLGISFLLILMLILYSFI 375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1767.5	89.9	539	5	Q25971	Q25971 plasmodium
2	1767.5	89.9	539	5	Q25973	Q25973 plasmodium
3	1763.5	89.7	400	5	Q03999	Q03999 plasmodium
4	1763.5	89.7	539	5	Q25966	Q25966 plasmodium
5	1763.5	89.7	539	5	Q25976	Q25976 plasmodium
6	1763.5	89.7	539	5	Q25984	Q25984 plasmodium
7	1763.5	89.7	539	5	Q9TYG1	Q9tyg1 plasmodium
8	1760.5	89.5	539	5	Q25972	Q25972 plasmodium
9	1760.5	89.5	539	5	Q25981	Q25981 plasmodium
10	1113.5	56.6	651	5	Q25924	Q25924 plasmodium
11	1110.5	56.5	569	5	Q25969	Q25969 plasmodium
12	1110.5	56.5	569	5	Q25974	Q25974 plasmodium
13	1110.5	56.5	569	5	Q25975	Q25975 plasmodium
14	1110.5	56.5	569	5	Q25977	Q25977 plasmodium
15	1110.5	56.5	569	5	Q25979	Q25979 plasmodium
16	1110.5	56.5	1694	5	Q9TZT5	Q9tzt5 plasmodium
17	1110.5	56.5	1694	5	Q9NHX1	Q9nhx1 plasmodium
18	1110.5	56.5	1704	5	Q9TZT4	Q9tzt4 plasmodium
19	1109.5	56.4	569	5	Q25983	Q25983 plasmodium
20	1097.5	55.8	569	5	Q25967	Q25967 plasmodium
21	1091.5	55.5	569	5	Q25970	Q25970 plasmodium
22	1091.5	55.5	569	5	Q25980	Q25980 plasmodium
23	1091.5	55.5	569	5	Q25982	Q25982 plasmodium
24	1091	55.5	570	5	Q25968	Q25968 plasmodium
25	1091	55.5	570	5	Q9TYG2	Q9tyg2 plasmodium
26	1090.5	55.4	569	5	Q25978	Q25978 plasmodium
27	1080	54.9	1720	5	Q25922	Q25922 plasmodium
28	1076	54.7	652	5	Q25923	Q25923 plasmodium
29	1015	51.6	373	5	Q25724	Q25724 plasmodium
30	1012.5	51.5	372	5	Q25725	Q25725 plasmodium
31	1012	51.4	373	5	Q25721	Q25721 plasmodium
32	1009.5	51.3	372	5	O43997	O43997 plasmodium
33	1009	51.3	372	5	Q25726	Q25726 plasmodium
34	1009	51.3	373	5	O43995	O43995 plasmodium
35	1005	51.1	373	5	Q25722	Q25722 plasmodium
36	1004	51.0	373	5	Q25723	Q25723 plasmodium
37	993.5	50.5	372	5	Q25717	Q25717 plasmodium
38	993.5	50.5	372	5	Q25719	Q25719 plasmodium
39	993.5	50.5	372	5	Q25720	Q25720 plasmodium
40	993	50.5	373	5	Q25727	Q25727 plasmodium
41	991	50.4	373	5	O43996	O43996 plasmodium
42	988.5	50.3	372	5	Q25718	Q25718 plasmodium
43	988	50.2	373	5	Q25728	Q25728 plasmodium
44	969.5	49.3	219	5	Q9UB87	Q9ub87 plasmodium
45	716	36.4	599	5	Q9NCN2	Q9ncn2 plasmodium

ALIGNMENTS

RESULT 1
Q25971
ID Q25971 PRELIMINARY; PRT; 539 AA.
AC Q25971;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13359; BAA02620.1; -.
DR InterPro; IPR000561; -.
DR InterPro; IPR001245; -.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1 1
SQ SEQUENCE 539 AA; 61017 MW; 5CA9C651BB62B5E1 CRC64;

Query Match 89.9%; Score 1767.5; DB 5; Length 539;
Best Local Similarity 91.8%; Pred. No. 6.8e-89;
Matches 346; Conservative 7; Mismatches 21; Indels 3; Gaps 2;

```
Qy      1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283
```

```
Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy    181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
          ||||||||||   |||| ||||||||||||||||||||||||||||||||||||||||||
Db    344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy    239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
          ||||| |::|||||| | : : : :     |    |||||||||||||||||||||||
Db    404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 462

Qy    299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
          ||||||||||||||||||||||||||||||||||||||:|||||||||   ||||||||
Db    463 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy    359 LGISFLLILMLILYSFI 375
          |||||||||||||||||
Db    523 LGISFLLILMLILYSFI 539


RESULT   2
Q25973
ID   Q25973         PRELIMINARY;   PRT;   539 AA.
AC   Q25973;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13361; BAA02622.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   539 AA;  61047 MW;  105C2E02974FB398 CRC64;


  Query Match             89.9%;  Score 1767.5;  DB 5;  Length 539;
  Best Local Similarity   91.8%;  Pred. No. 6.8e-89;
  Matches  346;  Conservative    7;  Mismatches   21;  Indels    3;  Gaps    2;

Qy      1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy    181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
          ||||||||||   |||| ||||||||||||||||||||||||||||||||||||||||||
Db    344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy    239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
          ||||| |::|||||| | : : : :     |    |||||||||||||||||||||||
Db    404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 462

Qy    299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
          ||||||||||||||||||||||||||||||||||||||:|||||||||   ||||||||
Db    463 GSKCVENPNPTCNENNGGCDADAKCTEECSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy    359 LGISFLLILMLILYSFI 375
          |||||||||||||||||
Db    523 LGISFLLILMLILYSFI 539


RESULT   3
Q03999
ID   Q03999      PRELIMINARY;      PRT;   400 AA.
AC   Q03999;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MEROZOITE SURFACE ANTIGEN PRECURSOR 1 (PMMSA) (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Blackman M.J., Ling I.T., Nicholls S.C., Holder A.A.;
RL   Submitted (XXX-1991) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: COULD BE ATTACHED TO THE MEMBRANE BY A GPI-
CC       ANCHOR.
CC   -!- MISCELLANEOUS: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF
CC       83K, 42K & 19K ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
DR   EMBL; M64681; AAA29709.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   NON_TER       1      1
FT   SIGNAL       <1     24       POTENTIAL.
FT   CHAIN        25    400       83K MEROZOITE SURFACE ANTIGEN.
FT   CHAIN        25    286       42K MEROZOITE SURFACE ANTIGEN.
FT   CHAIN       287    400       19K MEROZOITE SURFACE ANTIGEN.
FT   TRANSMEM    383    400       MEMBRANE ANCHOR.
SQ   SEQUENCE   400 AA;  45824 MW;  537F075058626AC2 CRC64;


  Query Match             89.7%;  Score 1763.5;  DB 5;  Length 400;
  Best Local Similarity   91.5%;  Pred. No. 8.1e-89;
  Matches  345;  Conservative    7;  Mismatches   22;  Indels    3;  Gaps    2;

Qy      1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     25 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 84

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 144

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 204

Qy    181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
          ||||||||||   |||| ||||||||||||||||||||||||||||||||||||||||||
Db    205 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 264

Qy    239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
          ||||| |::|||||| | : : : :     |    |||||||||||||||||||||||
Db    265 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 323

Qy    299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
```

```
          | ||||||||||||||||||||||||||||||||||||||:|||||||||   ||||||||
Db    324 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 383

Qy    359 LGISFLLILMLILYSFI 375
          |||||||||||||||||
Db    384 LGISFLLILMLILYSFI 400


RESULT   4
Q25966
ID   Q25966         PRELIMINARY;   PRT;   539 AA.
AC   Q25966;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13357; BAA02618.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   539 AA;  61144 MW;  7B7DE90C1D0ACDE7 CRC64;


  Query Match             89.7%;  Score 1763.5;  DB 5;  Length 539;
  Best Local Similarity   91.5%;  Pred. No. 1.1e-88;
  Matches  345;  Conservative    7;  Mismatches   22;  Indels    3;  Gaps    2;

Qy      1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy    181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
          ||||||||||   |||| ||||||||||||||||||||||||||||||||||||||||||
Db    344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy    239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
          ||||| |::|||||  | : : : :     |    ||||||||||||||||||||||||
Db    404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 462

Qy    299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
          | ||||||||||||||||||||||||||||||||||:|||||||||   ||||||||
Db    463 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy    359 LGISFLLILMLILYSFI 375
          |||||||||||||||||
Db    523 LGISFLLILMLILYSFI 539


RESULT   5
Q25976
ID   Q25976         PRELIMINARY;   PRT;   539 AA.
AC   Q25976;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13362; BAA02623.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   539 AA;  61045 MW;  8BF9C64322E9A778 CRC64;


  Query Match             89.7%;  Score 1763.5;  DB 5;  Length 539;
  Best Local Similarity   91.5%;  Pred. No. 1.1e-88;
  Matches  345;  Conservative    7;  Mismatches   22;  Indels    3;  Gaps    2;

Qy      1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy    181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
          ||||||||||   |||| ||||||||||||||||||||||||||||||||||||||||||
Db    344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy    239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
          ||||| |::|||||  | : : : :     |    ||||||||||||||||||||||||
Db    404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 462

Qy    299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
          | ||||||||||||||||||||||||||||||||||:|||||||||   ||||||||
Db    463 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy    359 LGISFLLILMLILYSFI 375
          |||||||||||||||||
Db    523 LGISFLLILMLILYSFI 539


RESULT   6
Q25984
ID   Q25984         PRELIMINARY;   PRT;   539 AA.
AC   Q25984;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13356; BAA02617.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   539 AA;  61114 MW;  3788015F3127CB9E CRC64;


  Query Match             89.7%;  Score 1763.5;  DB 5;  Length 539;
  Best Local Similarity   91.5%;  Pred. No. 1.1e-88;
  Matches  345;  Conservative    7;  Mismatches   22;  Indels    3;  Gaps    2;

Qy       1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
           ||||||| |||||||||||||||||||||||||||||||||||||||||||| |||||
Db     164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy     181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
           ||||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||
Db     344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy     239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
           ||||| |::|||||  | : : : :     |    ||||||||||||||||||||||
Db     404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 462

Qy     299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
           | |||||||||||||||||||||||||||||||||:|||||||||   ||||||||
Db     463 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy     359 LGISFLLILMLILYSFI 375
           |||||||||||||||||
Db     523 LGISFLLILMLILYSFI 539


RESULT   7
Q9TYG1
ID   Q9TYG1      PRELIMINARY;      PRT;   539 AA.
AC   Q9TYG1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13358; BAA02619.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   539 AA;  61075 MW;  C70C2E100EC4A101 CRC64;


  Query Match             89.7%;  Score 1763.5;  DB 5;  Length 539;
  Best Local Similarity   91.5%;  Pred. No. 1.1e-88;
  Matches  345;  Conservative    7;  Mismatches   22;  Indels    3;  Gaps    2;

Qy       1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
           ||||||| |||||||||||||||||||||||||||||||||||||||||||| |||||
Db     164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy     181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
           ||||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||
Db     344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy     239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
           ||||| |::|||||  | : : : :     |    ||||||||||||||||||||||
Db     404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 462

Qy     299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
           | |||||||||||||||||||||||||||||||||:|||||||||   ||||||||
Db     463 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy     359 LGISFLLILMLILYSFI 375
           |||||||||||||||||
Db     523 LGISFLLILMLILYSFI 539


RESULT   8
Q25972
ID   Q25972         PRELIMINARY;   PRT;   539 AA.
AC   Q25972;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13360; BAA02621.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   539 AA;  61076 MW;  72186F1412C57D1C CRC64;


  Query Match             89.5%;  Score 1760.5;  DB 5;  Length 539;
  Best Local Similarity   91.2%;  Pred. No. 1.6e-88;
```

```
  Matches  344;  Conservative    8;  Mismatches   22;  Indels    3;  Gaps    2;

Qy       1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy     181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
           ||||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||||
Db     344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy     239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
           ||||| |::|||||||  | : : : :     |    :|||||||||||||||||||||||
Db     404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQE 462

Qy     299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
           | ||||||||||||||||||||||||||||||||||:|||||||||   ||||||||
Db     463 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy     359 LGISFLLILMLILYSFI 375
           |||||||||||||||||
Db     523 LGISFLLILMLILYSFI 539


RESULT   9
Q25981
ID   Q25981         PRELIMINARY;   PRT;   539 AA.
AC   Q25981;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13363; BAA02624.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001245; -.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   539 AA;  61046 MW;  3EED87473EE87B65 CRC64;


  Query Match             89.5%;  Score 1760.5;  DB 5;  Length 539;
  Best Local Similarity   91.2%;  Pred. No. 1.6e-88;
  Matches  344;  Conservative    8;  Mismatches   22;  Indels    3;  Gaps    2;

Qy       1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     164 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 223

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     224 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 283

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     284 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 343

Qy     181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
           ||||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||||
Db     344 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 403

Qy     239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
           ||||| |::|||||||  | : : : :     |    :|||||||||||||||||||||||
Db     404 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQE 462

Qy     299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
           | ||||||||||||||||||||||||||||||||||:|||||||||   ||||||||
Db     463 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 522

Qy     359 LGISFLLILMLILYSFI 375
           |||||||||||||||||
Db     523 LGISFLLILMLILYSFI 539


RESULT  10
Q25924
ID   Q25924      PRELIMINARY;      PRT;   651 AA.
AC   Q25924;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MEROZOITE SURFACE ANTIGEN 1 (FRAGMENT).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RO-71;
RX   MEDLINE=92275047; PubMed=1592091;
RA   Olafsson P., Matile H., Certa U.;
RT   "Plasmodium falciparum: the repetitive MSA-1 surface protein of the
RT   RO-71 isolate is recognized by mouse antibody against the
RT   nonrepetitive repeat block of RO-33.";
RL   Exp. Parasitol. 74:381-389(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RO-71;
RX   MEDLINE=95354793; PubMed=7628566;
RA   Tolle R., Bujard H., Cooper J.A.;
RT   "Plasmodium falciparum: variations within the C-terminal region of
RT   merozoite surface antigen-1.";
RL   Exp. Parasitol. 81:47-54(1995).
DR   EMBL; Z35329; CAA84558.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite.
FT   NON_TER       1      1
SQ   SEQUENCE   651 AA;  74134 MW;  AA2137E699255150 CRC64;


  Query Match             56.6%;  Score 1113.5;  DB 5;  Length 651;
  Best Local Similarity   57.0%;  Pred. No. 2.9e-53;
  Matches  221;  Conservative   59;  Mismatches   85;  Indels   23;  Gaps    4;

Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
           ||||  ||||:|:|||||||||||||||:| |: |||:|: |||||   ||: | :||||
Db     266 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLES 325

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
           ||: :| ::|: |:::| :| || |::  | || |||:|:: || || : : ||: :
Db     326 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 385

Qy     129 KYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVNDKI 169
           ||| ||:|||| |                    |:| :  |:|||| ||||||  : :||
Db     386 KYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKI 445
```

```
Qy    170 DLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHN 227
          | ::|:|:||: :   ||  ::|:| :|  || | ||:  || :|:|  |  |  |
Db    446 DDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDDTKK 505

Qy    228 NLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDERE 287
          ::| | ||||:| :|   :|:| |:: |  : : : :     |    ||||||||||||
Db    506 DMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPQNSGCFRHLDERE 563

Qy    288 ECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPL 347
          |||||||||||| ||||||||||||||||||||||||||||||||||||:|||||||||
Db    564 ECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPL 623

Qy    348 SMVIFCSSSNFLGISFLLILMLILYSFI 375
          |||||||||||||||||||||||||||
Db    624 FDGIFCSSSNFLGISFLLILMLILYSFI 651


RESULT 11
Q25969
ID   Q25969         PRELIMINARY;   PRT;   569 AA.
AC   Q25969;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13346; BAA02607.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64485 MW;  A26C45914AF01396 CRC64;

  Query Match           56.5%;  Score 1110.5;  DB 5;  Length 569;
  Best Local Similarity   56.7%;  Pred. No. 3.7e-53;
  Matches  220;  Conservative   60;  Mismatches   85;  Indels   23;  Gaps    4;

Qy      9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
          ||||  ||||:|:|||||||||||||:| |::|||:|: |||||   ||: | :||||
Db    184 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLES 243

Qy     69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
          ||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db    244 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 303

Qy    129 KYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVNDKI 169
          ||| ||:|||| |                  |:| :  |:|||| ||||||  : :||
Db    304 KYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKI 363

Qy    170 DLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHN 227
          | ::|:|:||: :   ||  ::|:| :|  || | ||:  ||  |:|  |  |  |
Db    364 DDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKK 423

Qy    228 NLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDERE 287
          ::| | ||||:| :|   :|:| |:: |  : : : :     |   :||||||||||||
Db    424 DMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDERE 481

Qy    288 ECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPL 347
          |||||||||||| ||||||||||||||||||||||||||||||||||||:|||||||||
Db    482 ECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPL 541

Qy    348 SMVIFCSSSNFLGISFLLILMLILYSFI 375
          |||||||||||||||||||||||||||
Db    542 FDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT 12
Q25974
ID   Q25974         PRELIMINARY;   PRT;   569 AA.
AC   Q25974;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13348; BAA02609.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64457 MW;  A2691FE569E80396 CRC64;

  Query Match           56.5%;  Score 1110.5;  DB 5;  Length 569;
  Best Local Similarity   56.7%;  Pred. No. 3.7e-53;
  Matches  220;  Conservative   60;  Mismatches   85;  Indels   23;  Gaps    4;

Qy      9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
          ||||  ||||:|:|||||||||||||:| |::|||:|: |||||   ||: | :||||
Db    184 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLES 243

Qy     69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
          ||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db    244 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 303

Qy    129 KYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVNDKI 169
          ||| ||:|||| |                  |:| :  |:|||| ||||||  : :||
Db    304 KYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKI 363

Qy    170 DLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHN 227
          | ::|:|:||: :   ||  ::|:| :|  || | ||:  ||  |:|  |  |  |
Db    364 DDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKK 423

Qy    228 NLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDERE 287
          ::| | ||||:| :|   :|:| |:: |  : : : :     |   :||||||||||||
Db    424 DMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDERE 481

Qy    288 ECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPL 347
          |||||||||||| ||||||||||||||||||||||||||||||||||||:|||||||||
Db    482 ECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPL 541

Qy    348 SMVIFCSSSNFLGISFLLILMLILYSFI 375
          |||||||||||||||||||||||||||
Db    542 FDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT 13
Q25975
ID   Q25975         PRELIMINARY;   PRT;   569 AA.
AC   Q25975;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
```

```
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13349; BAA02610.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64459 MW;  ED4284B2867C9703 CRC64;


  Query Match             56.5%;  Score 1110.5;  DB 5;  Length 569;
  Best Local Similarity   56.7%;  Pred. No. 3.7e-53;
  Matches  220;  Conservative   60;  Mismatches   85;  Indels   23;  Gaps    4;

Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
           ||||  ||||:|:|||||||||||||||:| |::|||:|: |||||   ||: | :||||
Db     184 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLES 243

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
           ||: :| ::|: |:::| :| || |::  | || |||:|:: || || : : ||: :
Db     244 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 303

Qy     129 KYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVNDKI 169
           ||| ||:|||| |                  |:| :  |:|||| |||||  : :||
Db     304 KYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKI 363

Qy     170 DLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHN 227
           | ::|:|:||: :   ||  ::|:| :|  || | ||:  ||  |:|  |  |  |
Db     364 DDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKK 423

Qy     228 NLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDERE 287
           ::| | ||||:| :|   :|:| |:: |  : : : :     |    :|||||||||||
Db     424 DMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDERE 481

Qy     288 ECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPL 347
           |||||||||||| |||||||||||||||||||||||||||||||||||||:|||||||||
Db     482 ECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPL 541

Qy     348 SMVIFCSSSNFLGISFLLILMLILYSFI 375
           ||||||||||||||||||||||||||||
Db     542 FDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT  14
Q25977
ID   Q25977         PRELIMINARY;   PRT;   569 AA.
AC   Q25977;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13350; BAA02611.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64485 MW;  A26C45914AF01396 CRC64;


  Query Match             56.5%;  Score 1110.5;  DB 5;  Length 569;
  Best Local Similarity   56.7%;  Pred. No. 3.7e-53;
  Matches  220;  Conservative   60;  Mismatches   85;  Indels   23;  Gaps    4;

Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
           ||||  ||||:|:|||||||||||||||:| |::|||:|: |||||   ||: | :||||
Db     184 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLES 243

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
           ||: :| ::|: |:::| :| || |::  | || |||:|:: || || : : ||: :
Db     244 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 303

Qy     129 KYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVNDKI 169
           ||| ||:|||| |                  |:| :  |:|||| |||||  : :||
Db     304 KYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKI 363

Qy     170 DLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHN 227
           | ::|:|:||: :   ||  ::|:| :|  || | ||:  ||  |:|  |  |  |
Db     364 DDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKK 423

Qy     228 NLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDERE 287
           ::| | ||||:| :|   :|:| |:: |  : : : :     |    :|||||||||||
Db     424 DMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDERE 481

Qy     288 ECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPL 347
           |||||||||||| |||||||||||||||||||||||||||||||||||||:|||||||||
Db     482 ECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPL 541

Qy     348 SMVIFCSSSNFLGISFLLILMLILYSFI 375
           ||||||||||||||||||||||||||||
Db     542 FDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT  15
Q25979
ID   Q25979         PRELIMINARY;   PRT;   569 AA.
AC   Q25979;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13352; BAA02613.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64457 MW;  A2691FE569E80396 CRC64;


  Query Match             56.5%;  Score 1110.5;  DB 5;  Length 569;
  Best Local Similarity   56.7%;  Pred. No. 3.7e-53;
  Matches  220;  Conservative   60;  Mismatches   85;  Indels   23;  Gaps    4;
```

```
Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
           ||||  ||||:|:||||||||||||||:| |::|||:|: |||||   ||: | :||||
Db     184 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLES 243

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
           ||: :| ::|: |:::| :| || |:::  | ||.|||:|:: || || : : ||: :
Db     244 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 303

Qy     129 KYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVNDKI 169
           ||| ||:|||| |                   |:| :  |:|||| |||||  : :||
Db     304 KYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKI 363

Qy     170 DLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHN 227
           | ::|:|:||: :   ||  ::|:| :|  || | ||:  ||  |:|  |  |  |
Db     364 DDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKK 423

Qy     228 NLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDERE 287
           ::| | ||||:| :|   :|:| |:: |  : : : :     |    :|||||||||||
Db     424 DMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDERE 481

Qy     288 ECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPL 347
           |||||||||||| ||||||||||||||||||||||||||||||||||||:|||||||||
Db     482 ECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPL 541

Qy     348 SMVIFCSSSNFLGISFLLILMLILYSFI 375
             ||||||||||||||||||||||||||
Db     542 FDGIFCSSSNFLGISFLLILMLILYSFI 569
```

Search completed: August 8, 2001, 12:33:46
Job time: 274 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:32:34 ; Search time 24.79 Seconds
(without alignments)
518.185 Million cell updates/sec

Title: US-09-500-376-4
Perfect score: 1967
Sequence: 1 AVTPSVIHNILSKIENEYEV.....SNFLGISFLLILMLILYSFI 375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1763.5	89.7	1630	1	MSP1_PLAFK	P04932 plasmodium
2	1763.5	89.7	1639	1	MSP1_PLAFW	P04933 plasmodium
3	1110	56.4	1726	1	MSP1_PLAFC	P04934 plasmodium
4	1109	56.4	1726	1	MSP1_PLAFP	P50495 plasmodium
5	1102.5	56.0	1682	1	MSP1_PLAF3	P19598 plasmodium
6	1091	55.5	1701	1	MSP1_PLAFF	P13819 plasmodium
7	1091	55.5	1701	1	MSP1_PLAFM	P08569 plasmodium
8	565.5	28.7	1772	1	MSP1_PLAYO	P13828 plasmodium
9	149.5	7.6	1162	1	BXEN_CLOBO	P46082 clostridium
10	146.5	7.4	1162	1	BXEN_CLOBU	Q06366 clostridium
11	137.5	7.0	1956	1	ATX1_PLAFA	Q04956 plasmodium
12	136.5	6.9	377	1	Y704_METJA	Q58115 methanococc
13	135	6.9	1150	1	IRR1_YEAST	P40541 saccharomyc
14	134	6.8	971	1	Y228_BORBU	O51246 borrelia bu
15	133	6.8	864	1	CHEA_BORBU	Q44737 borrelia bu
16	132	6.7	1030	1	Y018_MYCPN	P75093 mycoplasma
17	129.5	6.6	2198	1	YLJ2_CAEEL	P34367 caenorhabdi
18	129	6.6	989	1	SERA_PLAFG	P13823 plasmodium
19	129	6.6	1251	1	RBP2_PLAVB	Q00799 plasmodium
20	127.5	6.5	867	1	DPO1_RICPR	O05949 rickettsia
21	127	6.5	756	1	Y328_MYCGE	Q49419 mycoplasma
22	126	6.4	1169	1	EX5B_BORBU	O51578 borrelia bu
23	125	6.4	442	1	TIG_BUCAI	P57546 buchnera ap
24	125	6.4	655	1	YKDA_MYCCA	P45615 mycoplasma
25	125	6.4	748	1	Y875_METJA	Q58285 methanococc
26	125	6.4	1196	1	BXCN_CLOBO	P46081 clostridium
27	124.5	6.3	749	1	MAD1_YEAST	P40957 saccharomyc
28	124.5	6.3	896	1	RPOP_NEUCR	P33540 neurospora
29	124	6.3	3305	1	APLP_MANSE	Q25490 manduca sex
30	123	6.3	881	1	LHS1_YEAST	P36016 saccharomyc
31	123	6.3	944	1	NUF1_YEAST	P32380 saccharomyc
32	122	6.2	922	1	LCN2_LACLA	P37609 lactococcus
33	122	6.2	1038	1	CIN8_YEAST	P27895 saccharomyc
34	120.5	6.1	629	1	PR39_YEAST	P39682 saccharomyc
35	120	6.1	1901	1	YCF1_TOBAC	P12222 nicotiana t
36	119.5	6.1	396	1	RMAR_YEAST	P02381 saccharomyc
37	119.5	6.1	1875	1	MLP1_YEAST	Q02455 saccharomyc
38	118.5	6.0	540	1	YKZ6_YEAST	P36112 saccharomyc
39	118.5	6.0	1024	1	RPOB_PLAFA	P21421 plasmodium
40	118.5	6.0	1077	1	CTR9_YEAST	P89105 saccharomyc
41	118.5	6.0	2009	1	SEC7_YEAST	P11075 saccharomyc
42	118	6.0	703	1	HS83_PHANI	P51819 pharbitis n
43	118	6.0	1738	1	YCF1_EPIVI	Q00383 epifagus vi
44	117.5	6.0	569	1	CYSP_PLAFA	P25805 plasmodium
45	117.5	6.0	766	1	YS83_CAEEL	Q09622 caenorhabdi

ALIGNMENTS

RESULT 1
MSP1_PLAFK
ID MSP1_PLAFK STANDARD; PRT; 1630 AA.
AC P04932;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
RT Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RA Pan W., Tolle R., Bujard H.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC ---

CC ---
DR EMBL; X03371; CAA27070.1; -.
DR PIR; A25120; SAZQK1.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.
FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.
FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 755 755 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 759 759 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 835 835 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
FT   CARBOHYD    955    955       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1049   1049       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1156   1156       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1165   1165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1436   1436       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1517   1517       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1630 AA;  187289 MW;  ADBDEC3CE0A46322 CRC64;


  Query Match           89.7%;  Score 1763.5;  DB 1;  Length 1630;
  Best Local Similarity   91.5%;  Pred. No. 8e-89;
  Matches  345;  Conservative    7;  Mismatches   22;  Indels    3;  Gaps    2;

Qy       1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
           ||||||| |||||||||||||||||||||||||||||||||||||||||||| |||||
Db    1255 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 1314

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1315 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 1374

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1375 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 1434

Qy     181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
           ||||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||||
Db    1435 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 1494

Qy     239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
           ||||| |::|||||| | : : : : :     |    ||||||||||||||||||||||||
Db    1495 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 1553

Qy     299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
           | |||||||||||||||||||||||||||||||||:|||||||||   ||||||||
Db    1554 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 1613

Qy     359 LGISFLLILMLILYSFI 375
           |||||||||||||||||
Db    1614 LGISFLLILMLILYSFI 1630


RESULT   2
MSP1_PLAFW
ID   MSP1_PLAFW     STANDARD;      PRT;  1639 AA.
AC   P04933;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P195).
GN   MSP-1.
OS   Plasmodium falciparum (isolate Wellcome).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5848;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86014355; PubMed=2995820;
RA   Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA   Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA   Freeman R.R.;
RT   "Primary structure of the precursor to the three major surface
RT   antigens of Plasmodium falciparum merozoites.";
RL   Nature 317:270-273(1985).
RN   [2]
RP   REVISIONS.
RA   Holder A.A.;
RL   Submitted (MAR-1991) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X02919; CAA26676.1; -.
DR   PIR; A24594; A24594.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1639       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD   116    116       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   268    268       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   764    764       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   768    768       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   783    783       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   844    844       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   920    920       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   964    964       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1058   1058       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1165   1165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1174   1174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1445   1445       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1526   1526       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1639 AA;  187618 MW;  2C255B6616C87F6E CRC64;


  Query Match           89.7%;  Score 1763.5;  DB 1;  Length 1639;
  Best Local Similarity   91.5%;  Pred. No. 8e-89;
  Matches  345;  Conservative    7;  Mismatches   22;  Indels    3;  Gaps    2;

Qy       1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
           ||||||| |||||||||||||||||||||||||||||||||||||||||||| |||||
Db    1264 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 1323

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1324 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 1383

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1384 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 1443

Qy     181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
           ||||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||||
Db    1444 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 1503

Qy     239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
           ||||| |::|||||| | : : : : :     |    ||||||||||||||||||||||||
Db    1504 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 1562

Qy     299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
           | |||||||||||||||||||||||||||||||||:|||||||||   ||||||||
Db    1563 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 1622

Qy     359 LGISFLLILMLILYSFI 375
           |||||||||||||||||
Db    1623 LGISFLLILMLILYSFI 1639


RESULT   3
MSP1_PLAFC
ID   MSP1_PLAFC     STANDARD;      PRT;  1726 AA.
AC   P04934;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
```

```
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P195).
GN   MSP-1.
OS   Plasmodium falciparum (isolate Camp / Malaysia).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5835;
RN   [1]
RP   SEQUENCE OF 1-1103 FROM N.A.
RX   MEDLINE=86205236; PubMed=3517809;
RA   Weber J.L., Leininger W.M., Lyon J.A.;
RT   "Variation in the gene encoding a major merozoite surface antigen of
RT   the human malaria parasite Plasmodium falciparum.";
RL   Nucleic Acids Res. 14:3311-3323(1986).
RN   [2]
RP   SEQUENCE OF 1104-1726 FROM N.A.
RX   MEDLINE=88143999; PubMed=3278296;
RA   Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
RT   "Merozoite surface protein sequence from the Camp strain of the human
RT   malaria parasite Plasmodium falciparum.";
RL   Nucleic Acids Res. 16:1206-1206(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03831; CAA27446.1; -.
DR   PIR; A23386; SAZQGM.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1726       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    133    133       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    567    567       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    638    638       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    924    924       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    944    944       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    990    990       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1016   1016       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1114   1114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1221   1221       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1613   1613       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1726 AA;  196197 MW;  DD8AD45FA352BCF3 CRC64;

  Query Match           56.4%;  Score 1110;  DB 1;  Length 1726;
  Best Local Similarity  56.6%;  Pred. No. 2.9e-53;
  Matches  220;  Conservative   60;  Mismatches   85;  Indels   24;  Gaps    4;

Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
           ||||  ||||:|:|||||||||||||||:| |: |||:|: |||||   ||: | :|||| 
Db    1340 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLES 1399

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
           ||: :| ::|: |:::| :| || |:::  | || |||:|:|:: || || : : ||: : 
Db    1400 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 1459

Qy     129 KYKSDLDSIKKYI-------------------NDKQGENEKYLPFLNNIETLYKTVNDK 168
           ||| ||:|||| |                    |:| :  |:|||| ||||||  : :| 
Db    1460 KYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNK 1519

Qy     169 IDLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNH 226
           || ::|:|:||: :   ||  ::|:| :|  || | ||:  || :|:|  |  |  |
Db    1520 IDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDDTK 1579

Qy     227 NNLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDER 286
            ::| | ||||:| :|   :|:| |:: |  : : : :     |    :|||||||||| 
Db    1580 KDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDER 1637

Qy     287 EECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYP 346
           |||||||||||| ||||||||||||||||||||||||||||||||||||:||||||| 
Db    1638 EECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYP 1697

Qy     347 LSMVIFCSSSNFLGISFLLILMLILYSFI 375
           |   |||||||||||||||||||||||||| 
Db    1698 LFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   4
MSP1_PLAFP
ID   MSP1_PLAFP     STANDARD;      PRT;  1726 AA.
AC   P50495;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (GP195).
GN   MSP-1.
OS   Plasmodium falciparum (isolate Palo Alto / Uganda).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=57270;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89005525; PubMed=3049134;
RA   Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
RA   Siddiqui W.A.;
RT   "Plasmodium falciparum: gene structure and hydropathy profile of the
RT   major merozoite surface antigen (gp195) of the Uganda-Palo Alto
RT   isolate.";
RL   Exp. Parasitol. 67:1-11(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M37213; AAA29611.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1726       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    133    133       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    567    567       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    638    638       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    924    924       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    944    944       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    990    990       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1016   1016       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1114   1114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1221   1221       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD   1613   1613       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1726 AA;  196174 MW;  5B59CEEFA2F9A026 CRC64;


  Query Match             56.4%;  Score 1109;  DB 1;  Length 1726;
  Best Local Similarity   56.6%;  Pred. No. 3.3e-53;
  Matches  220;  Conservative   60;  Mismatches   85;  Indels   24;  Gaps    4;

Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
           ||||  ||||:|:|||||||||||||||:| |: |||:|: |||||   ||: | :||||
Db    1340 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLES 1399

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
           ||: :| ::|: |:::| :| || |::  | || |||:|:: || || : : ||: :
Db    1400 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 1459

Qy     129 KYKSDLDSIKKYI-------------------NDKQGENEKYLPFLNNIETLYKTVNDK 168
           ||| ||:|||| |                    |:| :  |:|||| |||||  : :|
Db    1460 KYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNK 1519

Qy     169 IDLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNH 226
           || ::|:|:||: :   ||  ::|:| :|  || | ||:  || :|:|  |  |  |
Db    1520 IDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFDAIKKLINDDTK 1579

Qy     227 NNLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDER 286
           ::| | ||||:| :|   :|:| |:: |  : : : :     |    :||||||||||
Db    1580 KDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDER 1637

Qy     287 EECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYP 346
           |||||||||||| ||||||||||||||||||||||||||||||||||||:||||||||
Db    1638 EECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYP 1697

Qy     347 LSMVIFCSSSNFLGISFLLILMLILYSFI 375
           |   |||||||||||||||||||||||||||
Db    1698 LFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   5
MSP1_PLAF3
ID   MSP1_PLAF3     STANDARD;      PRT;  1682 AA.
AC   P19598; Q25921;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P190).
GN   MSP-1.
OS   Plasmodium falciparum (isolate ro-33 / Ghana).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5834;
RN   [1]
RP   SEQUENCE OF 1-1061 FROM N.A.
RX   MEDLINE=88166657; PubMed=3327688;
RA   Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT   "A naturally occurring gene encoding the major surface antigen
RT   precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
RL   EMBO J. 6:4137-4142(1987).
RN   [2]
RP   SEQUENCE OF 1032-1682 FROM N.A.
RX   MEDLINE=95354793; PubMed=7628566;
RA   Tolle R., Bujard H., Cooper J.A.;
RT   "Plasmodium falciparum: variations within the C-terminal region of
RT   merozoite surface antigen-1.";
RL   Exp. Parasitol. 81:47-54(1995).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M35727; AAA29715.1; -.
DR   EMBL; Y00087; CAA68280.1; -.
DR   EMBL; Z35326; CAA84555.1; -.
DR   PIR; S06286; S06286.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1682       MEROZOITE SURFACE PROTEIN 1.
FT   TRANSMEM  1666   1682       MEMBRANE ANCHOR (POTENTIAL).
FT   CARBOHYD   233    233       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   462    462       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   528    528       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   599    599       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   785    785       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   881    881       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   901    901       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   947    947       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1071   1071       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1178   1178       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1569   1569       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1682 AA;  192462 MW;  C82A1E159948CAD6 CRC64;


  Query Match             56.0%;  Score 1102.5;  DB 1;  Length 1682;
  Best Local Similarity   56.7%;  Pred. No. 7.2e-53;
  Matches  220;  Conservative   58;  Mismatches   87;  Indels   23;  Gaps    4;

Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
           ||||  ||||:|:|||||||||||||||:| |: |||:|: |||||   ||: | :||||
Db    1297 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLES 1356

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
           ||: :| ::|: |:::| :| || |::  | || |||:|:: || || : : ||: :
Db    1357 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 1416

Qy     129 KYKSDLDSIKKYI-------------------NDKQGENEKYLPFLNNIETLYKTVNDKI 169
           ||| ||:|||| |                    |:| :  |:|||| |||||  : :||
Db    1417 KYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKI 1476

Qy     170 DLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHN 227
           | ::|:|:||: :   ||  ::|:| :|  || | ||:  ||   :|  |  |  |
Db    1477 DDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFEAIKKLINDDTKK 1536

Qy     228 NLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDERE 287
           ::| | ||||:| :|   :|:| |:: |  : : : :     |    ||||||||||||
Db    1537 DMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPQNSGCFRHLDERE 1594

Qy     288 ECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPL 347
           |||||||||||| ||||||||||||||||||||||||||||||||||||:|||||||||
Db    1595 ECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPL 1654

Qy     348 SMVIFCSSSNFLGISFLLILMLILYSFI 375
             |||||||||||||||||||||||||||
Db    1655 FDGIFCSSSNFLGISFLLILMLILYSFI 1682


RESULT   6
MSP1_PLAFF
ID   MSP1_PLAFF     STANDARD;      PRT;  1701 AA.
AC   P13819;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA).
GN   MSP-1.
```

```
OS   Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88142999; PubMed=2449612;
RA   Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
RA   Brown G.V., Anders R.F., Kemp D.J.;
RT   "Variation in the precursor to the major merozoite surface antigens
RT   of Plasmodium falciparum.";
RL   Mol. Biochem. Parasitol. 27:291-302(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19143; AAA29653.1; -.
DR   PIR; A54498; A54498.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1701       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    536    536       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    607    607       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    802    802       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    899    899       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    965    965       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    991    991       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1089   1089       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1196   1196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1588   1588       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1701 AA;  193719 MW;  3920B75E73D38552 CRC64;


  Query Match             55.5%;  Score 1091;  DB 1;  Length 1701;
  Best Local Similarity   55.8%;  Pred. No. 3.1e-52;
  Matches  217;  Conservative   61;  Mismatches   87;  Indels   24;  Gaps    4;

Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
           ||||. ||||:|:||||||||||||||:| |::|||:|: |||||   ||: | :|||| 
Db    1315 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLES 1374

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
           ||: :| ::|: |:::| :| || |::  | || |||:|:: || || : : ||: :
Db    1375 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 1434

Qy     129 KYKSDLDSIKKYI-------------------NDKQGENEKYLPFLNNIETLYKTVNDK 168
           ||| ||:|||| |                     |:| :  |:|||| |||||  : :|
Db    1435 KYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNK 1494

Qy     169 IDLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNH 226
           || ::|:|:||: :   ||  ::|:| :|  || | ||:  ||  |:|  |  |  |
Db    1495 IDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTK 1554

Qy     227 NNLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDER 286
            ::| | ||||:| :|   :|:| |:: |  : : : :     |    :||||||||||
Db    1555 KDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDER 1612

Qy     287 EECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYP 346
           |||||||||||| |||||||||||||||||| |||||||: |||||:||||||||
Db    1613 EECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDSYP 1672

Qy     347 LSMVIFCSSSNFLGISFLLILMLILYSFI 375
           |   |||||||||||||||||||||||||
Db    1673 LFDGIFCSSSNFLGISFLLILMLILYSFI 1701


RESULT   7
MSP1_PLAFM
ID   MSP1_PLAFM     STANDARD;      PRT;  1701 AA.
AC   P08569;
DT   01-AUG-1988 (Rel. 08, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P190).
GN   MSP-1.
OS   Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=70153;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88011243; PubMed=3079521;
RA   Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT   "Allelic dimorphism in a surface antigen gene of the malaria parasite
RT   Plasmodium falciparum.";
RL   J. Mol. Biol. 195:273-287(1987).
RN   [2]
RP   REVISIONS TO 1403; 1569 AND 1629.
RA   Tanabe K.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-115 FROM N.A.
RX   MEDLINE=86136024; PubMed=3004972;
RA   Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA   Stunnenberg H., Bujard H.;
RT   "Polymorphism of the precursor for the major surface antigens of
RT   Plasmodium falciparum merozoites: studies at the genetic level.";
RL   EMBO J. 4:3823-3829(1985).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X05624; CAA29112.1; -.
DR   PIR; A26868; A26868.
DR   PIR; B25120; B25120.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1701       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    536    536       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    607    607       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    802    802       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    899    899       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    965    965       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    991    991       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1089   1089       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1196   1196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1588   1588       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1701 AA;  193768 MW;  3FC2EC59AF96EA98 CRC64;


  Query Match             55.5%;  Score 1091;  DB 1;  Length 1701;
  Best Local Similarity   55.8%;  Pred. No. 3.1e-52;
  Matches  217;  Conservative   61;  Mismatches   87;  Indels   24;  Gaps    4;

Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
Db    1315 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLES 1374

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
Db    1375 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 1434

Qy     129 KYKSDLDSIKKYI--------------------NDKQGENEKYLPFLNNIETLYKTVNDK 168
Db    1435 KYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNK 1494

Qy     169 IDLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNH 226
Db    1495 IDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTK 1554

Qy     227 NNLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDER 286
Db    1555 KDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDER 1612

Qy     287 EECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYP 346
Db    1613 EECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDSYP 1672

Qy     347 LSMVIFCSSSNFLGISFLLILMLILYSFI 375
Db    1673 LFDGIFCSSSNFLGISFLLILMLILYSFI 1701


RESULT   8
MSP1_PLAYO
ID   MSP1_PLAYO     STANDARD;      PRT;  1772 AA.
AC   P13828;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (230 KDA).
GN   MSP-1.
OS   Plasmodium berghei yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5862;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90205979; PubMed=2320061;
RA   Lewis A.P.;
RT   "Sequence analysis upstream of the gene encoding the precursor to the
RT   major merozoite surface antigens of Plasmodium yoelii.";
RL   Mol. Biochem. Parasitol. 39:285-288(1990).
RN   [2]
RP   SEQUENCE OF 1093-1772 FROM N.A.
RC   STRAIN=17XL;
RX   MEDLINE=88124889; PubMed=2448778;
RA   Burns J.M. Jr., Daly T.M., Vaidya A.B., Long C.A.;
RT   "The 3' portion of the gene for a Plasmodium yoelii merozoite surface
RT   antigen encodes the epitope recognized by a protective monoclonal
RT   antibody.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:602-606(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03612; AAA29762.1; -.
DR   EMBL; J04668; AAA29702.1; -.
DR   PIR; A28121; A28121.
DR   PIR; A45532; A45532.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   1772       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   406    406       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   646    646       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   829    829       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1018   1018       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1090   1090       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1408   1408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1446   1446       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1541   1541       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1629   1629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1680   1680       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT  1521   1521       L -> V (IN REF. 2).
SQ   SEQUENCE   1772 AA;  197230 MW;  9A6291658EB0F45D CRC64;


  Query Match             28.7%;  Score 565.5;  DB 1;  Length 1772;
  Best Local Similarity   34.9%;  Pred. No. 1.2e-23;
  Matches  134;  Conservative   73;  Mismatches  144;  Indels   33;  Gaps    8;

Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
Db    1401 DILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQKRNDFLEVLSH 1460

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
Db    1461 ELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFNKMVE 1520

Qy     129 KYKSDLDSIKKYI-------ND-KQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
Db    1521 LYNTQLAAVKEQIATIEAETNDTNKEEKKKYIPILEDLKGLYETVIGQAEEYSEELQNRL 1580

Qy     181 LNYTYEKSNVEI--KELIYLKTIQDKLADF----KKNNNFVGIADLSTDYNHNNLLTKFL 234
Db    1581 DNYKNEKAEFEILTKNLEKYIQIDEKLDEFVEHAENNKHIASIA----------LNNLN 1629

Qy     235 STGMVFENLLKSILSNLLDWK----LARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECK 290
Db    1630 KSGLVGEGESKKILAKMLNMDGMDLLGVDPKHVCVDTRD---IPKNAGCFRDDNGTEEWR 1686

Qy     291 CLLNYKQ-EGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSM 349
Db    1687 CLLGYKKGEGNTCVENNNPTCDINNGGCDPTASCQNAESTENSKKIICTCKEPTPNAYYE 1746

Qy     350 VIFCSSSNFLGISFLLILMLILYS 373
Db    1747 GVFCSSSSFMGLSILLIITLIVFN 1770


RESULT   9
BXEN_CLOBO
ID   BXEN_CLOBO     STANDARD;      PRT;  1162 AA.
AC   P46082;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
```

```
DE   BOTULINUM NEUROTOXIN TYPE E, NONTOXIC COMPONENT.
GN   ENT-120.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MASHIKE;
RX   MEDLINE=93195515; PubMed=8450310;
RA   Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT   "The complete nucleotide sequence of the gene encoding the nontoxic
RT   component of Clostridium botulinum type E progenitor toxin.";
RL   J. Gen. Microbiol. 139:79-86(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not remqved. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12697; BAA02194.1; -.
DR   InterPro; IPR000395; -.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136856 MW;  96468EDDDAE0F39D CRC64;
```

```
  Query Match             7.6%;  Score 149.5;  DB 1;  Length 1162;
  Best Local Similarity   21.5%;  Pred. No. 0.31;
  Matches  105;  Conservative   70;  Mismatches  159;  Indels  155;  Gaps   26;

Qy      14 IENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLESDL--- 70
           : ::|:  |: |   : : | | |      : :   | || |:  |  | ::  |:|
Db     173 LTHKYDEFYVDPALELIKCLIKSL---YYLYGIKPNDNLNIPYRLRNEFNSLEYSELNMI 229

Qy      71 -------IPYKDLTSSNYVVKDPYKFLNKEKR-DKFLSSY-------NYIKDSIDT---- 111
                | || | :: |   | | |::  |  :|: : |       ||| :||
Db     230 DFLISGGIDYKLLNTNPYWFIDKY-FIDTSKNFEKYKNDYEIKIKNNNYIANSIKLYLEQ 288

Qy     112 --DINFANDV----LGYYK-----ILSEKYKSDL----------DSIKKY-IND-KQGE 147
             ||   |:    | |:      :: |:| : |           |  | | ||  | |:
Db     289 KFKIN-VKDIWELNLSYFSKEFQIMMPERYNNALNHYYRKEFYVIDYFKNYNINGFKNGQ 347

Qy     148 NEKYLP-------FLNNIETLYKTVND------KIDLFVIHLEAKVLNY----------T 184
           :  ||        :|  | :   :|       | :::   |:  | |:
Db     348 IKTKLPLSKYNKEIINKPELIVNLINQNNTVLMKSNIYGDGLKGTVDNFYSNYIIPYNLN 407

Qy     185 YEKS-------NVEIKELIYLKTIQDK-LADFKKN-NNFVGIADLS------------- 221
           || |       || |:|:  :  | |: :  ::|| : |: : :::
Db     408 YEHSINYFYLDNVNIEEIEKIPPINDEDIYPYRKNADTFIPVYNITKAKEINTTTPLPVN 467

Qy     222 ------TDYNHNNLLTKFLSTGMVFENLLKSILSNLLDW-KLARYVKHFTTPMRKKTMIQ 274
                | |  || : ||       :|: | |:| :|: :  :| |   |    |   :
Db     468 YLQAQMIDSNDINLSSDFLKVISSKGSLVYSFLNNTMDYLEFIKYDKPIDT---DKKYYK 524

Qy     275 QNSGCFRHLDEREECKCLLNYKQEGSKCVENPNP--------------TCNENNGGCDA 319
              ||            ||  : ::  |  |                 | ||   :
Db     525 WLKAIFR-----------NYSLDITETQEISNQFGDTKIIPWIGRALNILNTNNSFVEE 572

Qy     320 DAKCTEEDSGSNGKKITCQCTKPDSYPLSMV------------IFCSSSNFLGISFLLI 366
                   :  : ||    | |  | ||:            |:|  :||
Db     573 FKNLGPISLINKKENITIPKIKIDEIPSSMLNFSFKDLSENLFNIYC-KNNF-------Y 624

Qy     367 LMLILYSFI 375
           |  | |:|:
Db     625 LKKIYYNFL 633
```

```
RESULT  10
BXEN_CLOBU
ID   BXEN_CLOBU     STANDARD;      PRT;  1162 AA.
AC   Q06366;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE E, NONTOXIC COMPONENT.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=93360835; PubMed=8355622;
RA   Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
RA   Takeshi K., Ohyama T., Isogai E., Isogai H.;
RT   "Similarity in nucleotide sequence of the gene encoding nontoxic
RT   component of botulinum toxin produced by toxigenic Clostridium
RT   butyricum strain BL6340 and Clostridium botulinum type E strain
RT   Mashike.";
RL   Microbiol. Immunol. 37:395-398(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12739; BAA02231.1; -.
DR   InterPro; IPR000395; -.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136829 MW;  C86E9BE932DA78E4 CRC64;
```

```
  Query Match             7.4%;  Score 146.5;  DB 1;  Length 1162;
  Best Local Similarity   21.3%;  Pred. No. 0.45;
  Matches  106;  Conservative   69;  Mismatches  151;  Indels  171;  Gaps   28;

Qy      14 IENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLESDL--- 70
           : ::|:  |: |   : : | | |      : :   | || |:  |  | ::  |:|
Db     173 LTHKYDEFYVDPALELIKCLIKSL---YYLYGIKPNDNLNIPYRLRNEFNSLEYSELDMI 229

Qy      71 -------IPYKDLTSSNYVVKDPYKFLNKEKR-DKFLSSY-------NYIKDSIDT---- 111
                | || | :: |   | | |::  |  :|: : |       ||| :||
Db     230 DFLISGGIDYKLLNTNPYWFIDKY-FIDTSKNFEKYKNDYEIKIKNNNYIANSIKLYLEQ 288

Qy     112 --DINFANDV----LGYYK-----ILSEKYKSDL----------DSIKKY-IND-KQGE 147
             ||   |:    | |:      :: |:| : |           |  | | ||  | |:
Db     289 KFKIN-VKDIWELNLSYFSKEFQIMMPERYNNALNHYYRKEFYVIDYFKNYNINGFKNGQ 347

Qy     148 NEKYLP-------FLNNIETLYKTVND------KIDLFVIHLEAKV-------------- 180
           :  ||        :|  | :   :|       | : :   |:  |
Db     348 IKTKLPLSKYNKEIINKPELIVNLINQNNTVLMKSNFYGDGLKGNVDNFYSNYIIPYNLN 407

Qy     181 ----LNYTYEKSNVEIKELIYLKTIQDK-LADFKKN-NNFVGIADLS------------- 221
               :||:|   || |:|:  :  | |: :  ::|| : |: : :::
Db     408 YEHSINYSY-LDNVNIEEIEKIPPINDEDIYPYRKNADTFIPVYNITKAKEINTTTPLPV 466

Qy     222 -------TDYNHNNLLTKFLSTGMVFENLLKSILSNLLDW-KLARYVKHFTTPMRKKTMI 273
                  | |  || : ||       :|: | |:| :|: :  :| |   |    |   |
Db     467 NYLQAQMIDSNDINLSSDFLKVISSKGSLVYSFLNNTMDYLEFIKYDKPIDT---DKKYY 523

Qy     274 QQNSGCFRHLDEREECKCLLNYKQEGSKCVENPNP--------------TCNENNGGCD 318
           :     ||            ||  : ::  |  |                 | ||
```

```
Db    524 KWLKAIFR------------NYSLDITETQEISNQFGDTKIIPWIGRALNILNTNN---- 567

Qy    319 ADAKCTEEDSG-------SNGKKITCQCTKPDSYPLSMV-------------IFCSSSNF 358
              ||          :  : ||    | |  | ||:              |:|  :||
Db    568 ---SFVEEFKNLGPIFLINKKENITIPKIKIDEIPSSMLNFSFKDLSENLFNIYC-KNNF 623

Qy    359 LGISFLLILMLILYSFI 375
                |  | |:|:
Db    624 -------YLKKIYYNFL 633


RESULT  11
ATX1_PLAFA
ID   ATX1_PLAFA     STANDARD;      PRT;  1956 AA.
AC   Q04956;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROBABLE CATION-TRANSPORTING ATPASE 1 (EC 3.6.1.-).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T9/96;
RX   MEDLINE=93132070; PubMed=8421054;
RA   Krishna S., Cowan G., Meade J.C., Wells R.A., Stringer J.R.,
RA   Robson K.J.;
RT   "A family of cation ATPase-like molecules from Plasmodium
RT   falciparum.";
RL   J. Cell Biol. 120:385-398(1993).
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2
CC       ATPASES). SUBFAMILY V.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X65738; CAA46646.1; -.
DR   InterPro; IPR001757; -.
DR   Pfam; PF00122; E1-E2_ATPase; 4.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
KW   Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding.
FT   DOMAIN       1     35       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     36    58       POTENTIAL.
FT   DOMAIN       59    61       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     62    80       POTENTIAL.
FT   DOMAIN       81    407      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     408   427      POTENTIAL.
FT   DOMAIN       428   440      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     441   462      POTENTIAL.
FT   DOMAIN       463   1818     CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     1819  1837     POTENTIAL.
FT   DOMAIN       1838  1845     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     1846  1863     POTENTIAL.
FT   DOMAIN       1864  1881     CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     1882  1905     POTENTIAL.
FT   DOMAIN       1906  1928     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     1929  1952     POTENTIAL.
FT   DOMAIN       1953  1956     CYTOPLASMIC (POTENTIAL).
FT   MOD_RES      496   496      PHOSPHORYLATION (PROBABLE).
FT   METAL        1760  1760     MAGNESIUM (BY SIMILARITY).
FT   METAL        1764  1764     MAGNESIUM (BY SIMILARITY).
FT   DOMAIN       246   251      POLY-ASN.
FT   DOMAIN       252   256      POLY-LYS.
FT   DOMAIN       937   941      POLY-ASN.
FT   DOMAIN       1344  1347     POLY-LYS.
FT   DOMAIN       1363  1372     POLY-ASN.
FT   DOMAIN       1680  1684     POLY-ASN.
SQ   SEQUENCE   1956 AA;  230285 MW;  AE708AAE99009335 CRC64;


  Query Match             7.0%;  Score 137.5;  DB 1;  Length 1956;
  Best Local Similarity  22.0%;  Pred. No. 2.5;
  Matches   72;  Conservative   52;  Mismatches  107;  Indels   97;  Gaps   14;

Qy      5 SVIHNILSK---IENEYEV--LYL----------------KPLAGVYRSLKKQLENNVM 42
          ::|||:  |   : | |::  ::|                ||    : ::|   |: :
Db   1337 NLIHNLYKKKKYMNNYYDIDEVHLIGNNNNNNNKNNSKEKKLPLKNKMKHIRKNESNDNI 1396

Qy     43 TFNVNVKDILNSPFNKRENFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSY 102
          |||    : ::                 |  || :   ||   |    | |:|   | :
Db   1397 TFNTYTSNNIH---------------LSKYKYVHHKNYYYPDSCTNLRKKKNSLFYNLK 1440

Qy    103 NYIKDSIDTDINFANDVLGYYKILSEKY-----------KSDLDSIKKYINDKQGENEK 150
           ||               ||:   :||          | :|  ||      | |: |
Db   1441 KYI---------------YYE--KKKYLQHCLLKHDNYKKVELPRIKDINYSYQMESIK 1482

Qy    151 YLPFLNNIETLYKTVNDKIDLFVIHLEAKVLNYTYEKSNVEIKELIYLK-TIQDKLADFK 209
            |::::   :  |  :  ::|  :    |  | |: :  |  || | :| :|  ::
Db   1483 TRNFIHSLSEQFAFSNLILSFYIIKNDD---NNVYNKNYIYNKNYIYNKNSICNK--NYI 1537

Qy    210 KNNNFVGIADLSTDYNHNNLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRK 269
           | |:  | : :  || ||:  |        :|:|    | ||     :::| |:  :|
Db   1538 CNKNY--IYNKNNIYNKNNIYNK--------KNILTHAKSVLLSGSSKKFLKFFSNIIR- 1586

Qy    270 KTMIQQNSGCFRH-LDEREECKCLLNYK 296
                  || | |::  | :  ||
Db   1587 -----------RHKLKEKKNKKNIKRYK 1603


RESULT  12
Y704_METJA
ID   Y704_METJA     STANDARD;      PRT;   377 AA.
AC   Q58115;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ0704.
GN   MJ0704.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67517; AAB98705.1; -.
```

```
DR   TIGR; MJ0704; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     26     46       POTENTIAL.
FT   TRANSMEM     67     87       POTENTIAL.
FT   TRANSMEM    108    128       POTENTIAL.
FT   TRANSMEM    135    155       POTENTIAL.
SQ   SEQUENCE   377 AA;  44630 MW;  2746A0A96425C285 CRC64;


  Query Match             6.9%;  Score 136.5;  DB 1;  Length 377;
  Best Local Similarity   21.7%;  Pred. No. 0.44;
  Matches   48;  Conservative   42;  Mismatches   60;  Indels   71;  Gaps    9;

Qy      26 LAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLESDLIPYKDLTSSNYVVKD 85
           |  ||  ::|:::   :   :   :::|  |||
Db     194 LFAVYEIIEKRIKVGDVYAVIKCLNMINKNFNKY-------------------------- 227

Qy      86 PYKFLNKEKRDKFLSSYNYIKDSIDTD-----INFAND---------------VLGYYKI 125
            : |:  |||:|:|  :  |   :  :     :|  |                 :: :||
Db     228 -WMFIKDEKREKYLRDFLRILSKLRVEYRKNCLNRKNKPYSKKGLEAFKKTEFIVSFYKT 286

Qy     126 LSEKYKS-DLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDK--IDLFVIHLEAKVLN 182
           : || |: |::|| |::      :| |    |||   |:  |||  :::|| ||:  :
Db     287 IEEKIKTRDINSINKFL------SETY----NNISN-YEMFNDKTYLEIFVHHLKELLEK 335

Qy     183 YTYEKSNVEIKELIY----------LKTIQDKLADFKKNNN 213
           |  ||:  :: | |:          :    :|| : :  ||
Db     336 YEKEKNENKLDEEIFNELKDELEKLINKCNNKLRELESQNN 376


RESULT  13
IRR1_YEAST
ID   IRR1_YEAST     STANDARD;      PRT;  1150 AA.
AC   P40541;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   IRR1 PROTEIN.
GN   IRR1 OR YIL026C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=96090137; PubMed=7483852;
RA   Kurlandzka A., Rytka J., Gromadka R., Murawski M.;
RT   "A new essential gene located on Saccharomyces cerevisiae chromosome
RT   IX.";
RL   Yeast 11:885-890(1995).
CC   -!- FUNCTION: NOT KNOWN; MAY PARTICIPATE IN CELL-CELL AND/OR CELL-
CC       SUBSTRATUM INTERACTIONS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: SOME, TO S.POMBE SPAC17H9.20.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z46881; CAA86966.1; -.
DR   SGD; S0001288; IRR1.
FT   DOMAIN       44     48       POLY-GLU.
FT   DOMAIN       65     70       POLY-ASP.
SQ   SEQUENCE   1150 AA;  133008 MW;  89688EA09485AC28 CRC64;


  Query Match             6.9%;  Score 135;  DB 1;  Length 1150;
  Best Local Similarity   20.7%;  Pred. No. 1.9;
  Matches   89;  Conservative   73;  Mismatches  136;  Indels  132;  Gaps   22;

Qy       1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLE----NNVMTFNVNVKDILN--- 53
             |   | | |  ||:: :|:: : | :  ::   ||:|:    | : |  :::|  |
Db     733 AFMDSTIGN--SKVDMKYDI-FAKFIHHIHHFEKKELQEKFLNQIATLKIHLKKFLQEKM 789

Qy      54 SPFNKRENFKNVLESDLIPYKDLTSSNYVVKDPYKFLNK----------EKRDKFLSSY- 102
           | | |::           |||| | | :     ::||          |  :: |  :
Db     790 DPNNSRDD-----------YKDLTCSLYEL-----YINKLTILGRDYPIEVDEELLQLFL 833

Qy     103 -NYIK----------DSIDTDINFANDVLGYYKILS----EKYKSDLDSIKKYIN----D 143
            |::           ||   :|||   ||      |:    ||::  :: :: | |    |
Db     834 NNFVSRIPIMFQDFDDSTAQEINFKMLVL-----LATWNLEKWREIIEKVRDYENSISKD 888

Qy     144 KQGENEKYLPFLNNIETL----------YKTVND----------------KIDLFVIHL 176
           :   :     :  : ||          :: :|                  | :| : |
Db     889 LRSVWKPIAAIIGRLNTLVISLAATNETFENINSLFYLKWSACTSLMDIIVAIKIFELKL 948

Qy     177 EAKVLNYTY-----------EKSNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDY 224
           |    : |              | | :|  :||:::  |  | :       :||   :
Db     949 PADATTWRYSMSEQFPFYLHDNASKVLLKIFLYLESLFAKQVDVQLER----VADEDANL 1004

Qy     225 N-------HNNLLTKFLSTGMVFENLLKSILSNLLDWKLARYV---KHFTTPMRKKTM-- 272
           |       |: |:||   :  : |:|    |:|| : |  |   |    |: || :
Db    1005 NDLPETGFFENIETEFLLFTVKLKGLMK---LNILDERFASRVALNKEKLGPLFKKIVDD 1061

Qy     273 -IQQN--------SGCFRHLDEREECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKC 323
           | :|           :  :||:     | ::|     :: |   |: :     | :
Db    1062 TIMENPEPNKKNIQKAKSNQTQREKAPLQPNSERE----TDHANTENNDPDIPMTIDLEP 1117

Qy     324 TEEDSGSNGK 333
           || | :| :
Db    1118 IEESSQNNSE 1127


RESULT  14
Y228_BORBU
ID   Y228_BORBU     STANDARD;      PRT;   971 AA.
AC   O51246;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL PROTEIN BB0228.
GN   BB0228.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001133; AAC66621.1; -.
DR   TIGR; BB0228; -.
KW   Hypothetical protein.
SQ   SEQUENCE   971 AA;  112959 MW;  088A688D7BBC591A CRC64;


  Query Match             6.8%;  Score 134;  DB 1;  Length 971;
  Best Local Similarity   19.8%;  Pred. No. 1.8;
  Matches   74;  Conservative   60;  Mismatches  102;  Indels  138;  Gaps   17;

Qy      26 LAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLESDLIPYKDLTSSNYVVKD 85
           ::|:   ||: :      |: :::::    || : |||:: |:|   |:|   | : |:
Db     325 ISGINTDLKESI------FSFGLQNVVE---NKEKEFKNLVFSEL---KNLV-KNKIPKE 371

Qy      86 -------PYKFLNKEKRDK----------------------FLSSY--NYIKDSIDTDI 113
                 |:|  ||:: :                        :||  | | : ::  |
Db     372 LIKGILFGYEFALKEEKGQNFPIALMIKSFKGWLNGLHPIKTLQTSYYINEITNKLEKGI 431

Qy     114 NFANDVLGYYKILSEKY----------------------------------------KSD 133
            :  :::  | | :  |                                          | |
Db     432 YYFENLIEKYLIFNNHYTLISFIPSHDTEKEMEEEIEKKLMAREIEIKQNPEEFLQFKKD 491

Qy     134 LDSIKKYINDKQGENE-KYLPFLNNIETLYKTVNDKIDLFVIHLEAKVLNYTYEKSNV-- 190
           :  ||| | |  : :   || |  || | | :   :||  |  |  : :: :: :|:
Db     492 YNQFKKYQNKKDSKADIAKLPLL-KIEDLPKQIEKSLDLNEIK-ELNLHSFKFKSNNIFN 549

Qy     191 ----------EIKELIYLKTIQDKLADFKKNN-----------NFVGIADLSTDYNHN-- 227
                    | :: |||   :  | |    |           | :|  ::|  |: :
Db     550 VNLFFKLDFLEKEDYIYLSLFKRALQDLSTKNYSYININNKIQNTLGQINISESYDEDID 609

Qy     228 -NLLTKFLSTGMVFEN-------LLKSILSNL----------------LDWKLARYVK-H 262
            |:|  |  :   | |       |:| || |:                 |:|     | |
Db     610 GNILNSFNISFKSFNNKVKESFELIKEILININFHDYERLKEITLSLKNDFKSLLIPKGH 669

Qy     263 FTTPMRKKTMIQQN 276
           :| |: :: |
Db     670 LLAMLRSKSKLKLN 683


RESULT  15
CHEA_BORBU
ID   CHEA_BORBU     STANDARD;      PRT;   864 AA.
AC   Q44737; P70857; Q44877;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CHEMOTAXIS PROTEIN CHEA (EC 2.7.3.-).
GN   CHEA OR BB0669.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=212;
RX   MEDLINE=98438936; PubMed=9765799;
RA   Trueba G.A., Old I.G., Saint-Girons I., Johnson R.C.;
RT   "A cheA cheW operon in Borrelia burgdorferi, the agent of Lyme
RT   disease.";
RL   Res. Microbiol. 148:191-200(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
RN   [3]
RP   SEQUENCE OF 1-30 FROM N.A.
RC   STRAIN=212;
RX   MEDLINE=97144545; PubMed=8990312;
RA   Ge Y., Charon N.W.;
RT   "An unexpected flaA homolog is present and expressed in Borrelia
RT   burgdorferi.";
RL   J. Bacteriol. 179:552-556(1997).
CC   -!- FUNCTION: INVOLVED IN THE TRANSMISSION OF SENSORY SIGNALS FROM
CC       THE CHEMORECEPTORS TO THE FLAGELLAR MOTORS. CHEA IS
CC       AUTOPHOSPHORYLATED; IT CAN TRANSFER ITS PHOSPHATE GROUP TO EITHER
CC       CHEB OR CHEY (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC   -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC       KINASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U28962; AAB96835.1; -.
DR   EMBL; AE001168; AAC67024.1; -.
DR   EMBL; U62900; AAC44771.1; -.
DR   EMBL; X91907; CAA63002.1; -.
DR   TIGR; BB0669; -.
DR   InterPro; IPR000410; -.
DR   InterPro; IPR002545; -.
DR   InterPro; IPR002570; -.
DR   Pfam; PF01584; CheW; 1.
DR   Pfam; PF01627; Hpt; 1.
DR   Pfam; PF00512; signal; 1.
DR   PRINTS; PR00344; BCTRLSENSOR.
KW   Sensory transduction; Transferase; Kinase; Phosphorylation;
KW   Chemotaxis.
FT   MOD_RES      51     51       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CONFLICT     11     11       L -> I (IN REF. 1).
FT   CONFLICT    182    182       S -> H (IN REF. 1).
FT   CONFLICT    187    187       S -> G (IN REF. 1).
FT   CONFLICT    239    239       L -> S (IN REF. 1).
FT   CONFLICT    362    362       A -> S (IN REF. 1).
FT   CONFLICT    559    559       S -> P (IN REF. 1).
SQ   SEQUENCE   864 AA;  98352 MW;  C1111DE0877BE624 CRC64;


  Query Match             6.8%;  Score 133;  DB 1;  Length 864;
  Best Local Similarity   22.1%;  Pred. No. 1.8;
  Matches   73;  Conservative   53;  Mismatches  104;  Indels  100;  Gaps   14;

Qy       5 SVIHNILSKIENEYEVLYLK---------PLAGVYRSLKKQLENNVMTFNVNVKDILNSP 55
           ||:| |  | |   |  :||         : || :||          : |: ::  |
Db     200 SVLHTI-PKYEQIIEDKFLKRVDYYLIYSDIEGVKKSLD--------SLNL-IESYLVDE 249

Qy      56 FNKRENFKNVLESDLIPYKDL-TSSNYVVKDPYKFLNKEKRDKFLSSYN---------YI 105
           || :|  | : : ::   ||:   ||:|: | : |   |  |  |   |          ::
Db     250 FNVKEELKKLADEEI---KDVDLDSNFVLNDNFDFTEDEISDLLLEVENQKLFKVRLDFV 306

Qy     106 KDSIDTDINFANDVLGYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTV 165
           ||:    |:    :|   | | : :||  ||  : : ||  :   |    |  |   :::
Db     307 KDNPMATIS-GLQMLQALKSLGKIFKSIPDS-SELLADKFFDFVIYYLISNTSE---ESI 361

Qy     166 NDKIDL--FVIHLEAKVLNYTYEKSNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTD 223
```

```
          ||:|   | | | |         || :: |  ::  :|  | ||:| |    : :| :
Db    362 AKKINLPDVVSHFEIK---------NVNLESLKSVRLKEDDEAPFKENKNIKKNSPISVN 412

Qy    224 ------------------------------------YNHN--------------NLLT 231
                                                 || |              :|
Db    413 LIRIDSKKIDYILNLVSEAVISKSSYNQINSEMITLFYNFNYFYDYQESFQRNFLIDLKI 472

Qy    232 KFLSTGMVFENLLKSILSNLLDWKLARYVK 261
           |   |:  |: ::| :::|: :|: : :|
Db    473 VFKDAGLTLEDEIESHINSLMSFKMEKALK 502
```

Search completed: August 8, 2001, 12:32:38
Job time: 206 sec

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:30:54 ; Search time 43.63 Seconds
(without alignments)
654.721 Million cell updates/sec

Title: US-09-500-376-4
Perfect score: 1967
Sequence: 1 AVTPSVIHNILSKIENEYEV.....SNFLGISFLLILMLILYSFI 375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1914	97.3	1631	1	SAZQK1	major merozoite su
2	1901	96.6	1640	2	A24594	probable major sur
3	1763.5	89.7	400	2	A45545	major merozoite su
4	1763.5	89.7	1639	2	S05603	major merozoite su
5	1113.5	56.6	651	2	S47282	merozoite surface
6	1110	56.4	1726	1	SAZQGM	major merozoite su
7	1103	56.1	1726	2	A45948	major merozoite su
8	1091	55.5	1701	2	A54498	major merozoite su
9	1077	54.8	1701	2	A26868	major merozoite su
10	714	36.3	1726	2	A39401	merozoite surface
11	713	36.2	1751	2	A45604	major blood-stage
12	565.5	28.7	680	2	A28121	major merozoite su
13	565.5	28.7	1772	2	A45532	major merozoite su
14	529	26.9	1785	2	A45546	major merozoite su
15	165.5	8.4	1306	2	T28313	ORF MSV152 probabl
16	164.5	8.4	2269	2	T28677	rhoptry protein -
17	157	8.0	1127	2	T28317	ORF MSV156 hypothe
18	156	7.9	1714	2	E71609	Ser/Thr protein ki
19	153	7.8	1346	2	G71613	hypothetical prote
20	150.5	7.7	1162	2	A47708	progenitor toxin n
21	149.5	7.6	706	2	A57719	dihydropteroate sy
22	148	7.5	507	2	A71622	hypothetical prote
23	146.5	7.4	1162	2	I40817	botulinum toxin no
24	146	7.4	722	2	D71607	VPS45-like protein
25	145.5	7.4	2401	2	T28676	rhoptry protein -
26	145	7.4	960	2	S72284	DNA-directed RNA p
27	145	7.4	980	2	E71606	hypothetical prote
28	143	7.3	608	2	T28301	ORF MSV140 hypothe
29	142.5	7.2	2010	2	B71616	phosphatase (acid
30	142.5	7.2	2708	2	T09079	probable chloroqui
31	142.5	7.2	2819	2	T09080	probable chloroqui
32	142.5	7.2	3394	2	T18501	hypothetical prote
33	142	7.2	522	2	T18504	hypothetical prote
34	142	7.2	1619	2	T18499	hypothetical prote
35	141.5	7.2	3973	2	B71612	hypothetical prote
36	140	7.1	1119	2	T18491	hypothetical prote
37	139.5	7.1	1817	2	D71606	hypothetical prote
38	139.5	7.1	2013	2	C71610	probable membrane
39	139	7.1	439	2	T28196	hypothetical prote
40	138.5	7.0	1045	2	H82934	type I restriction
41	137.5	7.0	1984	2	A44396	P-type cation tran
42	137	7.0	504	2	D71615	hypothetical prote
43	137	7.0	558	2	C71609	hypothetical prote
44	136.5	6.9	377	2	H64387	hypothetical prote
45	136.5	6.9	1005	2	A64465	hypothetical prote

ALIGNMENTS

RESULT 1
SAZQK1
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000
C;Accession: A25120
R;Mackay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.
EMBO J. 4, 3823-3829, 1985
A;Title: Polymorphism of the precursor for the major surface antigens of Plasmodium f
A;Reference number: A91030; MUID:86136024
A;Accession: A25120
A;Molecule type: DNA
A;Residues: 1-1631 <MAC>
C;Comment: The merozoite stages of different strains have strain-specific surface ant
C;Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane pr
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1631/Product: major merozoite surface antigen #status predicted <MAT>
F;67-84/Region: 3-residue repeats (S-G-T/P)
F;1614-1631/Domain: membrane anchor #status predicted <MBN>
F;97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (Asn

Query Match 97.3%; Score 1914; DB 1; Length 1631;
Best Local Similarity 97.9%; Pred. No. 9.6e-97;
Matches 369; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

```
Qy       1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    1255 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 1314

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1315 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 1374

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1375 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 1434

Qy     181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
           |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db    1435 LNYTYEKSNVEVKIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 1494

Qy     239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1495 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 1554

Qy     299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
           ||||||||||||||||||||||||||||||||||:|||||||||   ||||||||
Db    1555 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 1614
```

```
Qy    359 LGISFLLILMLILYSFI 375
          |||||||||||||||||
Db   1615 LGISFLLILMLILYSFI 1631
```

RESULT 2
A24594
probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodium
C;Species: Plasmodium falciparum
C;Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jun-2000
C;Accession: A24594
R;Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls,
Nature 317, 270-273, 1985
A;Title: Primary structure of the precursor to the three major surface antigens of Plasm
A;Reference number: A24594; MUID:86014355
A;Accession: A24594
A;Molecule type: DNA
A;Residues: 1-1640 <HOL>
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

Query Match 96.6%; Score 1901; DB 2; Length 1640;
Best Local Similarity 97.1%; Pred. No. 4.9e-96;
Matches 366; Conservative 3; Mismatches 6; Indels 2; Gaps 1;

```
Qy      1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
          ||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1264 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 1323

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 1383

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1384 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 1443

Qy    181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
          |||||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1444 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 1503

Qy    239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
          ||||||||:|||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db   1504 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 1563

Qy    299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
          ||||||| ||||||||||||||||||||||||||||:||||| |||||||||||||||||
Db   1564 GSKCVENSNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLSMVIFCSSSNF 1623

Qy    359 LGISFLLILMLILYSFI 375
          |||||||||||||||||
Db   1624 LGISFLLILMLILYSFI 1640

RESULT 3
A45545
major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments)
C;Species: Plasmodium falciparum
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
C;Accession: A45545
R;Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.
Mol. Biochem. Parasitol. 49, 29-33, 1991
A;Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protein-1
A;Reference number: A45545; MUID:92131048
A;Accession: A45545
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-400 <BLA>
A;Note: sequence extracted from NCBI backbone (NCBIN:77612, NCBIP:77621)
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

Query Match 89.7%; Score 1763.5; DB 2; Length 400;
Best Local Similarity 91.5%; Pred. No. 3e-89;
Matches 345; Conservative 7; Mismatches 22; Indels 3; Gaps 2;

```
Qy   1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   25 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 84

Qy   61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 144

Qy   121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   145 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 204

Qy   181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
          ||||||||||| |||| ||||||||||||||||||||||||||||||||||||||||||||
Db   205 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 264

Qy   239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
          ||||| |::|||||| | : : : : | ||||||||||||||||||||||||
Db   265 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 323

Qy   299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
          | ||||||||||||||||||||||||||||||||||:||||||||| ||||||||
Db   324 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 383

Qy   359 LGISFLLILMLILYSFI 375
          |||||||||||||||||
Db   384 LGISFLLILMLILYSFI 400

RESULT 4
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N;Alternate names: gp195 surface antigen
C;Species: Plasmodium falciparum
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C;Accession: S05603; S04850
R;Myler, P.J.
submitted to the EMBL Data Library, April 1989
A;Reference number: S05603
A;Accession: S05603
A;Molecule type: mRNA
A;Residues: 1-1639 <MYL>
A;Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
R;Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989
A;Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Pl
A;Reference number: S04850; MUID:89345116
A;Accession: S04850
A;Molecule type: mRNA
A;Residues: 1504-1639 <MYL2>
A;Cross-references: EMBL:X15063
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 89.7%; Score 1763.5; DB 2; Length 1639;
Best Local Similarity 91.5%; Pred. No. 1.5e-88;
Matches 345; Conservative 7; Mismatches 22; Indels 3; Gaps 2;

```
Qy      1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1264 AVTPSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 1323

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 1383
```

```
Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1384 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 1443

Qy    181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
          |||||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1444 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 1503

Qy    239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
          ||||| |::|||||| | : : : :     |    |||||||||||||||||||||||||
Db   1504 VFENLAKTVLSNLLDGNL-QGMLNISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQE 1562

Qy    299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
          | ||||||||||||||||||||||||||||||||||:|||||||||   ||||||||
Db   1563 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNF 1622

Qy    359 LGISFLLILMLILYSFI 375
          |||||||||||||||||
Db   1623 LGISFLLILMLILYSFI 1639


RESULT   5
S47282
merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71) (f
C;Species: Plasmodium falciparum
A;Variety: strain RO-71
C;Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 09-Jun-2000
C;Accession: S47282
R;Tolle, R.; Bujard, H.; Cooper, J.A.
submitted to the EMBL Data Library, July 1994
A;Description: Plasmodium falciparum: recombination within the C-terminal region of merc
A;Reference number: S47282
A;Accession: S47282
A;Molecule type: DNA
A;Residues: 1-651 <TOL>
A;Cross-references: EMBL:Z35329; NID:g535257; PIDN:CAA84558.1; PID:g535258
A;Experimental source: strain RO-71
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen


  Query Match            56.6%;  Score 1113.5;  DB 2;  Length 651;
  Best Local Similarity  57.0%;  Pred. No. 1.3e-53;
  Matches  221;  Conservative   59;  Mismatches   85;  Indels   23;  Gaps    4;

Qy      9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
          ||||  ||||:|:|||||||||||||:| |: |||:|: |||||   ||: | :||||
Db    266 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLES 325

Qy     69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
          ||: :| ::|: |:::| :| || |::  | || |||:|:: || || : : ||: :
Db    326 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 385

Qy    129 KYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVNDKI 169
          ||| ||:|||| |                  |:| :  |:|||| ||||||  : :||
Db    386 KYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKI 445

Qy    170 DLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHN 227
          | ::|:|:||: :   ||  ::|:| :|  || | ||:  || :|:|  |  |  |
Db    446 DDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDDTKK 505

Qy    228 NLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDERE 287
          ::| | ||||:| :|   :|:| |:: |  : : : :     |    ||||||||||||
Db    506 DMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPQNSGCFRHLDERE 563

Qy    288 ECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPL 347
          |||||||||||| ||||||||||||||||||||||||||||||||||:|||||||||
Db    564 ECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPL 623

Qy    348 SMVIFCSSSNFLGISFLLILMLILYSFI 375
            |||||||||||||||||||||||||||
Db    624 FDGIFCSSSNFLGISFLLILMLILYSFI 651


RESULT   6
SAZQGM
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N;Alternate names: 195K glycoprotein
C;Species: Plasmodium falciparum
C;Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
C;Accession: A23386; S06361
R;Weber, J.L.; Leininger, W.M.; Lyon, J.A.
Nucleic Acids Res. 14, 3311-3323, 1986
A;Title: Variation in the gene encoding a major merozoite surface antigen of the huma
A;Reference number: A23386; MUID:86205236
A;Accession: A23386
A;Molecule type: DNA
A;Residues: 1-1104 <WEB1>
A;Cross-references: EMBL:X03831
R;Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.
Nucleic Acids Res. 16, 1206, 1988
A;Title: Merozoite surface protein sequence from the Camp strain of the human malaria
A;Reference number: S06361; MUID:88143999
A;Accession: S06361
A;Molecule type: DNA
A;Residues: 1104-1726 <WEB2>
A;Cross-references: EMBL:X03831
C;Comment: The merozoite stages of different strains have strain-specific surface ant
C;Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1726/Product: major merozoite surface antigen #status predicted <MAT>
F;67-87,91-96,100-105,109-120/Region: 3-residue repeats (S-G-T)
F;757-765/Region: 3-residue repeats (T-E-E)
F;133,272,501,567,638,827,839,924,944,990,1016,1114,1221,1613,1658/Binding site: carb


  Query Match            56.4%;  Score 1110;  DB 1;  Length 1726;
  Best Local Similarity  56.6%;  Pred. No. 6.1e-53;
  Matches  220;  Conservative   60;  Mismatches   85;  Indels   24;  Gaps    4;

Qy      9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
          ||||  ||||:|:|||||||||||||:| |: |||:|: |||||   ||: | :||||
Db   1340 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLES 1399

Qy     69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
          ||: :| ::|: |:::| :| || |::  | || |||:|:: || || : : ||: :
Db   1400 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 1459

Qy    129 KYKSDLDSIKKYI------------------NDKQGENEKYLPFLNNIETLYKTVNDK 168
          ||| ||:|||| |                  |:| :  |:|||| ||||||  : :|
Db   1460 KYKDDLESIKKVIKEEKEFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNK 1519

Qy    169 IDLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNH 226
          || ::|:|:||: :   ||  ::|:| :|  || | ||:  || :|:|  |  |  |
Db   1520 IDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDDTK 1579

Qy    227 NNLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDER 286
           ::| | ||||:| :|   :|:| |:: |  : : : :     |    :|||||||||||
Db   1580 KDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDER 1637

Qy    287 EECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYP 346
          |||||||||||| ||||||||||||||||||||||||||||||||||:|||||||||
Db   1638 EECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYP 1697

Qy    347 LSMVIFCSSSNFLGISFLLILMLILYSFI 375
          |   |||||||||||||||||||||||||||
Db   1698 LFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   7
A45948
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
```

```
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jun-2000
C;Accession: A45948
R;Chang, S.P.; Kramer, K.J.; Yamaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.
Exp. Parasitol. 67, 1-11, 1988
A;Title: Plasmodium falciparum: gene structure and hydropathy profile of the major meroz
A;Reference number: A45948; MUID:89005525
A;Accession: A45948
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1726 <CHA>
A;Cross-references: GB:M37213
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match            56.1%;  Score 1103;  DB 2;  Length 1726;
  Best Local Similarity  56.3%;  Pred. No. 1.5e-52;
  Matches  219;  Conservative   60;  Mismatches   86;  Indels   24;  Gaps    4;

Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
Db    1340 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLES 1399

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
Db    1400 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 1459

Qy     129 KYKSDLDSIKKYI--------------------NDKQGENEKYLPFLNNIETLYKTVNDK 168
Db    1460 KYKDDLESIKKVIKEEKEKFPSSPPTTTPSPVKTDEQKKESKFLPFLTNIETLYNNLVNK 1519

Qy     169 IDLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNH 226
Db    1520 NDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDDTK 1579

Qy     227 NNLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDER 286
Db    1580 KDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDER 1637

Qy     287 EECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYP 346
Db    1638 EECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYP 1697

Qy     347 LSMVIFCSSSNFLGISFLLILMLILYSFI 375
Db    1698 LFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   8
A54498
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (is
C;Species: Plasmodium falciparum
C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C;Accession: A54498
R;Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.;
Mol. Biochem. Parasitol. 27, 291-302, 1988
A;Title: Variation in the precursor to the major merozoite surface antigens of Plasmodiu
A;Reference number: A54498; MUID:88142999
A;Accession: A54498
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1701 <PET>
A;Cross-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match            55.5%;  Score 1091;  DB 2;  Length 1701;
  Best Local Similarity  55.8%;  Pred. No. 6.5e-52;
  Matches  217;  Conservative   61;  Mismatches   87;  Indels   24;  Gaps    4;

Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
Db    1315 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLES 1374

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
Db    1375 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 1434

Qy     129 KYKSDLDSIKKYI--------------------NDKQGENEKYLPFLNNIETLYKTVNDK 168
Db    1435 KYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNK 1494

Qy     169 IDLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNH 226
Db    1495 IDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTK 1554

Qy     227 NNLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDER 286
Db    1555 KDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLDER 1612

Qy     287 EECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYP 346
Db    1613 EECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDSYP 1672

Qy     347 LSMVIFCSSSNFLGISFLLILMLILYSFI 375
Db    1673 LFDGIFCSSSNFLGISFLLILMLILYSFI 1701


RESULT   9
A26868
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
C;Accession: A26868
R;Tanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.
J. Mol. Biol. 195, 273-287, 1987
A;Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmod
A;Reference number: A26868; MUID:88011243
A;Accession: A26868
A;Molecule type: DNA
A;Residues: 1-1701 <TAN>
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1701/Product: major merozoite surface antigen #status predicted <MAT>


  Query Match            54.8%;  Score 1077;  DB 2;  Length 1701;
  Best Local Similarity  55.5%;  Pred. No. 3.8e-51;
  Matches  217;  Conservative   60;  Mismatches   86;  Indels   28;  Gaps    5;

Qy       9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
Db    1315 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLES 1374

Qy      69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
Db    1375 DLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKVLA 1434

Qy     129 KYKSDLDSIKKYI--------------------NDKQGENEKYLPFLNNIETLYKTVNDK 168
Db    1435 KYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNK 1494

Qy     169 IDLFVIHLEAKVLNYTYEK--SNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNH 226
Db    1495 IDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTK 1554

Qy     227 NNLLTKFLSTGMV--FENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLD 284
Db    1555 KDMLGKLLSTGLVQIFPN---TIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLD 1610

Qy     285 EREECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDS 344
Db    1611 EREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDS 1670
```

```
Qy    345 YPLSMVIFCSSSNFLGISFLLILMLILYSFI 375
          |||   |||||||||||||||||||||||||
Db   1671 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 1701
```

RESULT 10
A39401
merozoite surface antigen 1 precursor - Plasmodium vivax
C;Species: Plasmodium vivax
C;Date: 28-Feb-1992 #sequence_revision 28-Feb-1992 #text_change 24-Nov-1999
C;Accession: A39401
R;del Portillo, H.A.; Longacre, S.; Khouri, E.; David, P.H.
Proc. Natl. Acad. Sci. U.S.A. 88, 4030-4034, 1991
A;Title: Primary structure of the merozoite surface antigen 1 of Plasmodium vivax reveal
A;Reference number: A39401; MUID:91219506
A;Accession: A39401
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1726 <DEL>
A;Cross-references: GB:M60807
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

```
  Query Match             36.3%;  Score 714;  DB 2;  Length 1726;
  Best Local Similarity   39.6%;  Pred. No. 2.2e-31;
  Matches  155;  Conservative   67;  Mismatches  119;  Indels   50;  Gaps    7;

Qy     16 NEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLESDLIPYKD 75
          ::|:|:|||||||:|: :||||||:|  || |: |:|:|   ||  |  || ||| |:|
Db   1346 SDYDVVYLKPLAGMYKKIKKQLENHVNAFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKY 1405

Qy     76 LTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSEKYKSDL- 134
            |::|||||| |: ||: | | || ||  ||| |:  ||| : ||  : | ||: |
Db   1406 SPSGEYIIKDPYKLLDLEKKKKLLGSYKYIGASIDKDLGTANDGVNYYNKMGELYKTHLT 1465

Qy    135 ----------------DSIKKYIND------------KQGENEKYLPFLNNIETLYKTV 165
                           | |||  :|            |: | ||||||||:::  |:::
Db   1466 AVNEEVKKVEADIKAEDDKIKKIGSDSTKTTEKTQSMAKKAELEKYLPFLNSLQKEYESL 1525

Qy    166 NDKIDLFVIHLEAKVLNYTYEKSNVEI--KELIYLKTIQDKLADFKKNNNFVGIADLSTD 223
           |:: :  :|:  : |   ||   ||  |:|     : :|| ::||       ::   :
Db   1526 VSKVNTYTDNLKKVINNCQLEKKEAEITVKKLQDYNKMDEKLEEYKK-------SEKKNE 1578

Qy    224 YNHNNLLTKFLSTGMVFENLLKSILSNLLDWK---LARYVKHFTTPMRKKTMIQQNSGCF 280
            : || | : : :: ||  | ||| ||: :   |    :|        | :  |: |:
Db   1579 VKSSGLLEKLMKSKLIKENESKEILSQLLNVQTQLLTMSSEHTCI----DTNVPDNAACY 1634

Qy    281 RHLDEREECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCT 340
          |:||  || :||| :|:|| |||   | || :|||||  :|:|   ||     || |:||
Db   1635 RYLDGMEEWRCLLTFKEEGGKCVPGSNVTCKDNNGGCAPEAECKMTDS----NKIVCKCT 1690

Qy    341 KPDSYPLSMVIFCSSSNFLGISFLLILMLIL 371
          |  | ||   :|||||:|| :||||:::| |
Db   1691 KEGSEPLFEGVFCSSSSFLSLSFLLLMLLFL 1721
```

RESULT 11
A45604
major blood-stage surface antigen Pv200 - Plasmodium vivax
C;Species: Plasmodium vivax
C;Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
C;Accession: A45604
R;Gibson, H.L.; Tucker, J.E.; Kaslow, D.C.; Krettli, A.U.; Collins, W.E.; Kiefer, M.C.;
Mol. Biochem. Parasitol. 50, 325-333, 1992
A;Title: Structure and expression of the gene for Pv200, a major blood-stage surface ant
A;Reference number: A45604; MUID:92158013
A;Accession: A45604
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1751 <GIB>
A;Cross-references: GB:M75674; NID:g160608; PID:g457336
A;Note: sequence extracted from NCBI backbone (NCBIN:83591, NCBIP:83592)
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

```
  Query Match             36.2%;  Score 713;  DB 2;  Length 1751;
  Best Local Similarity   39.1%;  Pred. No. 2.6e-31;
  Matches  153;  Conservative   71;  Mismatches  117;  Indels   50;  Gaps    7;

Qy     16 NEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLESDLIPYKD 75
          ::|:|:|||||||:|:|:::|||||:|  || |: |:|:|   ||  |  || ||| |:|
Db   1371 SDYDVVYLKPLAGMYKTIKKQLENHVNAFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKY 1430

Qy     76 LTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSEKYKSDLD 135
             :|  |::||||| |: ||: | : || ||  ||| |:  ||| : ||  : | ||: ||
Db   1431 SSSGEYIIKDPYKLLDLEKKKKLIGSYKYIGASIDMDLATANDGVTYYNKMGELYKTHLD 1490

Qy    136 SIK---KYIND--------------------------KQGENEKYLPFLNNIETLYKTV 165
             :|   | : |                         |: | ||||||||:::  |:::
Db   1491 GVKTEIKKVEDDIKKQDEELKKLGNVNSQDSKKNEFIAKKAELEKYLPFLNSLQKEYESL 1550

Qy    166 NDKIDLFVIHLEAKVLNYTYEKSNVEI--KELIYLKTIQDKLADFKKNNNFVGIADLSTD 223
             |:: :  :|:  : |   ||   ||  |:|     : :|| ::||       ::   :
Db   1551 VSKVNTYTDNLKKVINNCQLEKKEAEITVKKLQDYNKMDEKLEEYKK-------SEKKNE 1603

Qy    224 YNHNNLLTKFLSTGMVFENLLKSILSNLLDWK---LARYVKHFTTPMRKKTMIQQNSGCF 280
            : || | : : :: ||  | ||| ||: :   |    :|        | :  |: |:
Db   1604 VKSSGLLEKLMKSKLIKENESKEILSQLLNVQTQLLTMSSEHTCI----DTNVPDNAACY 1659

Qy    281 RHLDEREECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCT 340
          |:||  || :||| :|:|| |||   | || :|||||  :|:|   ||     || |:||
Db   1660 RYLDGTEEWRCLLTFKEEGGKCVPASNVTCKDNNGGCAPEAECKMTDS----NKIVCKCT 1715

Qy    341 KPDSYPLSMVIFCSSSNFLGISFLLILMLIL 371
          |  | ||   :|||||:|| :||||:::| |
Db   1716 KEGSEPLFEGVFCSSSSFLSLSFLLLMLLFL 1746
```

RESULT 12
A28121
major merozoite surface antigen - Plasmodium yoelii (fragment)
C;Species: Plasmodium yoelii
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 26-Aug-1999
C;Accession: A28121
R;Burns Jr., J.M.; Daly, T.M.; Vaidya, A.B.; Long, C.A.
Proc. Natl. Acad. Sci. U.S.A. 85, 602-606, 1988
A;Title: The 3' portion of the gene for a Plasmodium yoelii merozoite surface antigen
A;Reference number: A28121; MUID:88124889
A;Accession: A28121
A;Molecule type: DNA
A;Residues: 1-680 <BUR>
A;Cross-references: GB:J03612; NID:g160678; PID:g160679
A;Experimental source: strain 17XL
A;Note: the authors translated the codon GTA for residue 429 as Leu
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

```
  Query Match             28.7%;  Score 565.5;  DB 2;  Length 680;
  Best Local Similarity   34.9%;  Pred. No. 9.4e-24;
  Matches  134;  Conservative   73;  Mismatches  144;  Indels   33;  Gaps    8;

Qy      9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
          :|||:  ||   :| | |   |:|||| :     |   :: : ||:   || :|  ||
Db    309 DILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQKRNDFLEVLSH 368

Qy     69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
          :|  :|||::: ||:::||: |: :|:|| : :  |    |: ||    | : ::  : |
Db    369 ELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFNKMVE 428

Qy    129 KYKSDLDSIKKYI-------ND-KQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           | : | ::|: |          ||  : | :||:| | ::: ||:||  : : :   |: ::
Db    429 VYNTQLAAVKEQIATIEAETNDTNKEEKKKYIPILEDLKGLYETVIGQAEEYSEELQNRL 488
```

```
Qy    181 LNYTYEKSNVEI--KELIYLKTIQDKLADF----KKNNNFVGIADLSTDYNHNNLLTKFL 234
          ||  ||:  ||  | |     | :|| :|    : | :   ||           |
Db    489 DNYKNEKAEFEILTKNLEKYIQIDEKLDEFVEHAENNKHIASIA----------LNNLN 537

Qy    235 STGMVFENLLKSILSNLLDWK----LARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECK 290
          :|:| |   | ||: :|:     |    ||      |    | :|:|||| :  || :
Db    538 KSGLVGEGESKKILAKMLNMDGMDLLGVDPKHVCVDTRD---IPKNAGCFRDDNGTEEWR 594

Qy    291 CLLNYKQ-EGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSM 349
          ||| ||: ||: |||| |||: |||||  | |   :| | ||| | | :|
Db    595 CLLGYKKGEGNTCVENNNPTCDINNGGCDPTASCQNAESTENSKKIICTCKEPTPNAYYE 654

Qy    350 VIFCSSSNFLGISFLLILMLILYS 373
          :|||||:|:|:| |||: ||:::
Db    655 GVFCSSSSFMGLSILLIITLIVFN 678


RESULT  13
A45532
major merozoite surface antigen precursor - Plasmodium yoelii
C;Species: Plasmodium yoelii
C;Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 26-Aug-1999
C;Accession: A45532; A45531
R;Lewis, A.P.
Mol. Biochem. Parasitol. 36, 271-282, 1989
A;Title: Cloning and analysis of the gene encoding the 230-kilodalton merozoite surface
A;Reference number: A45532; MUID:90014981
A;Accession: A45532
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1772 <LEW>
A;Cross-references: GB:J04668; NID:g160492; PID:g160493
R;Daly, T.M.; Burns Jr., J.M.; Long, C.A.
Mol. Biochem. Parasitol. 36, 283-285, 1989
A;Title: Precursor to the major merozoite surface antigen of Plasmodium yoelii: cloning
A;Reference number: A45531; MUID:90014982
A;Accession: A45531
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 454-1094 <DAL>
A;Cross-references: GB:J03975; NID:g160081; PID:g160082
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match             28.7%;  Score 565.5;  DB 2;  Length 1772;
  Best Local Similarity   34.9%;  Pred. No. 2.8e-23;
  Matches  134;  Conservative   73;  Mismatches  144;  Indels   33;  Gaps    8;

Qy      9 NILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLES 68
          :|||:  ||   :| | |   |:|||| :     |   :: : ||:   || :|  ||
Db   1401 DILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQKRNDFLEVLSH 1460

Qy     69 DLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKILSE 128
          :|  :|||::: ||:::||: |: :|:|| : :  |    |: ||    | : ::  : |
Db   1461 ELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFNKMVE 1520

Qy    129 KYKSDLDSIKKYI-------ND-KQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          | : | ::|: |       ||  : | :||:| | ::: ||:||  : : :   |: ::
Db   1521 LYNTQLAAVKEQIATIEAETNDTNKEEKKKYIPILEDLKGLYETVIGQAEEYSEELQNRL 1580

Qy    181 LNYTYEKSNVEI--KELIYLKTIQDKLADF----KKNNNFVGIADLSTDYNHNNLLTKFL 234
          ||  ||:  ||  | |     | :|| :|    : | :   ||           |
Db   1581 DNYKNEKAEFEILTKNLEKYIQIDEKLDEFVEHAENNKHIASIA----------LNNLN 1629

Qy    235 STGMVFENLLKSILSNLLDWK----LARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECK 290
          :|:| |   | ||: :|:     |    ||      |    | :|:|||| :  || :
Db   1630 KSGLVGEGESKKILAKMLNMDGMDLLGVDPKHVCVDTRD---IPKNAGCFRDDNGTEEWR 1686

Qy    291 CLLNYKQ-EGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSM 349
          ||| ||: ||: |||| |||: |||||  | |   :|  | ||| | | :|
Db   1687 CLLGYKKGEGNTCVENNNPTCDINNGGCDPTASCQNAESTENSKKIICTCKEPTPNAYYE 1746

Qy    350 VIFCSSSNFLGISFLLILMLILYS 373
          :|||||:|:|:| |||: ||:::
Db   1747 GVFCSSSSFMGLSILLIITLIVFN 1770


RESULT  14
A45546
major merozoite surface antigen precursor - Plasmodium chabaudi chabaudi
C;Species: Plasmodium chabaudi chabaudi
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
C;Accession: A45546
R;Deleersnijder, W.; Hendrix, D.; Bendahman, N.; Hanegreefs, J.; Brijs, L.; Hamers-Ca
Mol. Biochem. Parasitol. 43, 231-244, 1990
A;Title: Molecular cloning and sequence analysis of the gene encoding the major meroz
A;Reference number: A45546; MUID:91218805
A;Accession: A45546
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1785 <DEL>
A;Cross-references: GB:M34947; NID:g160597; PID:g160598
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match             26.9%;  Score 529;  DB 2;  Length 1785;
  Best Local Similarity   31.5%;  Pred. No. 2.8e-21;
  Matches  122;  Conservative   79;  Mismatches  158;  Indels   28;  Gaps    7;

Qy      3 TPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENF 62
          |   : :||   ::| | :| | |   |:| || :          ::   ||    || :|
Db   1409 TDEEVKDILDAFKSENEYIYTKSLGNTYKSFKKHMLKEFSMIKEDIMTGLNYKLEKRNDF 1468

Qy     63 KNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGY 122
           :||  :|  :||: :: :|||:||: |: :|:|| : :  |    :  ||  | | : :
Db   1469 LDVLSYELALFKDINTNKFVVKNPYQLLDNDKKDKQMINLKYAIKGVTEDIETATDGIEF 1528

Qy    123 YKILSEKYKSDLDSIKKYI-----NDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLE 177
          :  : | ||  |::: : |          | :|| |   ::: ||:|: :  : |   |:
Db   1529 FNKMIELYKPQLNAVNEQIAAIGTEPTDAEKKKYAPIFEDLKGLYETILNGAEEFSELLQ 1588

Qy    178 AKVLNYTYEKSNVEIKELIYLKT---IQDKLADF----KKNNNFVGIADLSTDYNHNNLL 230
          |: ||  ||:  :|  :  |:|   | :|| ||    :|| :   ||           |
Db   1589 HKLENYKIEKAGFDIL-MANLETYIRIDEKLEDFVESAEKNKHIASIA----------L 1636

Qy    231 TKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECK 290
              :|:| |   | ||: :|:      :   :  :   |    |:||||: |  || :
Db   1637 NNLNKSGLVTEGESKKILAKMLNMDAMDLLGIGSNHVCISTSTPDNAGCFRYDDGTEEWR 1696

Qy    291 CLLNYKQ--EGSKCVENPNPTCNENNGGCDADAKCTEEDSGSN--GKKITCQCTKPDSYP 346
          ||| :|:  :|::|| :  | || |||||| :| | | ::      ||| | | :|:
Db   1697 CLLGFKKDDDGNRCVADDAPVCNNNNGGCDKNADCREVENTDRDPSKKIVCTCKEPNPNA 1756

Qy    347 LSMVIFCSSSNFLGISFLLILMLILYS 373
              :||||| |:|:| |||: ||:::
Db   1757 YYAGVFCSSSGFMGLSILLIITLIVFN 1783


RESULT  15
T28313
ORF MSV152 probable core protein P4a homolog (vaccinia A10L) - Melanoplus sanguinipes
C;Species: Melanoplus sanguinipes entomopoxvirus
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T28313
R;Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A;Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A;Reference number: Z20484; MUID:99102612
A;Accession: T28313
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1306 <AFO>
A;Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97675.1; PID:g4049715
```

C;Genetics:
A;Note: MSV152

Query Match 8.4%; Score 165.5; DB 2; Length 1306;
Best Local Similarity 21.1%; Pred. No. 0.12;
Matches 72; Conservative 63; Mismatches 111; Indels 95; Gaps 13;

```
Qy       8 HNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNV-- 65
           ||: : | |: |:                 | : :  || |  : |   |   :|   :
Db      89 HNLQNIINNKTEI---------------LPDELSFFNYNTANYLRYRFMYDDNSSGLRG 132

Qy      66 ------LESDLIPYKDLTSSNYVVKDPYKF---LNKEKRDKFLSSYNYIKDSIDTDINFA 116
                 |   | | |  |::|:||    :   :|::| | ::: : :|      |
Db     133 GMLIDKTEDGFISYND-ESASYIVKKMENYISIINEDKYDFYVTYHAFI------DYFLE 185

Qy     117 NDVLGYYKILSEKYKSDL-----DSIKKYINDKQGENE---------------------- 149
           || | |  |: || |:|      :  |||:|:    |:
Db     186 NDKLSYGDIIDEKSKNDKMYNKDEDYKKYLNEHTVSNDESKKIRKKIKYYLKFFDTILHK 245

Qy     150 ------KYLPFLNNIETLYKTVNDKIDLFVIHLEAK-VLNYTYEKSNVEIKELIYLKTIQ 202
                :|:|    :: : |  : |: | ::  :|: :||  ' | :: ||   :: |
Db     246 DIISFYEYIPLTKILDYIDKDTDIKLILDILITKAEPLLN---ENNSTEIIAYLFNKKQS 302

Qy     203 DKLADFKK----NNN---------------FVGIADLSTDYNHNNLLTKFLS--TGMVF 240
            :| : ||    |||                 : | ::::| |   :|  | |    ::
Db     303 LRLYENKKYYNNNNNALSKFVSNKLLDIQRRIINIMEINSDNNPTEILITFNSKYQSLII 362

Qy     241 ENLLKSIL--SNLLDWKLARYVKHFTTPMRKKTMIQQNSGC 279
           |:  : : :| ::  |  |  :  :|  || : : |  |
Db     363 NNIYNAFIFFTNYINNYLIVYEFNLDSPNVKKFIKKFNDLC 403
```

Search completed: August 8, 2001, 12:30:58
Job time: 106 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:58 ; Search time 34.41 Seconds
(without alignments)
224.393 Million cell updates/sec

Title: US-09-500-376-4
Perfect score: 1967
Sequence: 1 AVTPSVIHNILSKIENEYEV.....SNFLGISFLLILMLILYSFI 375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	272	13.8	53	1	US-08-290-919-3	Sequence 3, Appli
2	253	12.9	53	1	US-08-290-919-4	Sequence 4, Appli
3	232	11.8	106	1	US-08-290-919-11	Sequence 11, Appl
4	200	10.2	48	1	US-08-290-919-1	Sequence 1, Appli
5	200	10.2	48	1	US-08-290-919-12	Sequence 12, Appl
6	197	10.0	48	1	US-08-290-919-2	Sequence 2, Appli
7	127	6.5	984	1	US-08-257-073-3	Sequence 3, Appli
8	127	6.5	984	2	US-08-184-009-120	Sequence 120, App
9	127	6.5	984	2	US-08-458-356-120	Sequence 120, App
10	117.5	6.0	2391	2	US-08-446-855A-2	Sequence 2, Appli
11	117.5	6.0	2391	4	US-09-150-741-2	Sequence 2, Appli
12	117	5.9	3135	1	US-08-323-170B-2	Sequence 2, Appli
13	114	5.8	1786	4	US-08-973-462-8	Sequence 8, Appli
14	113	5.7	455	2	US-08-392-625-21	Sequence 21, Appl
15	113	5.7	455	2	US-08-466-961A-21	Sequence 21, Appl
16	113	5.7	455	2	US-08-645-193B-23	Sequence 23, Appl
17	113	5.7	978	2	US-08-415-593-43	Sequence 43, Appl
18	112.5	5.7	1147	3	US-08-470-260-5	Sequence 5, Appli
19	112.5	5.7	1147	3	US-08-471-491-5	Sequence 5, Appli
20	112.5	5.7	1147	4	US-08-466-662-5	Sequence 5, Appli
21	112.5	5.7	2710	1	US-08-480-604A-6	Sequence 6, Appli
22	112.5	5.7	2710	2	US-08-405-496A-6	Sequence 6, Appli
23	112.5	5.7	3289	2	US-08-477-451-2	Sequence 2, Appli
24	112	5.7	912	2	US-08-951-871-2	Sequence 2, Appli
25	111.5	5.7	599	2	US-08-910-551B-2	Sequence 2, Appli
26	109.5	5.6	1435	2	US-08-568-459A-4	Sequence 4, Appli
27	109.5	5.6	1435	2	US-08-487-826B-4	Sequence 4, Appli
28	108.5	5.5	720	2	US-08-840-236-1	Sequence 1, Appli
29	108.5	5.5	720	2	US-08-505-448A-1	Sequence 1, Appli
30	108.5	5.5	990	2	US-08-392-625-20	Sequence 20, Appl
31	108.5	5.5	990	2	US-08-466-961A-20	Sequence 20, Appl
32	107.5	5.5	990	2	US-08-645-193B-15	Sequence 15, Appl
33	107	5.4	1264	1	US-07-789-915A-6	Sequence 6, Appli
34	107	5.4	1264	1	US-08-005-002C-6	Sequence 6, Appli
35	107	5.4	1264	1	US-08-487-203A-6	Sequence 6, Appli
36	105.5	5.4	615	1	US-08-484-105-6	Sequence 6, Appli
37	105.5	5.4	615	1	US-08-484-106-6	Sequence 6, Appli
38	105	5.3	934	3	US-08-929-329-6	Sequence 6, Appli
39	105	5.3	1579	3	US-08-755-587-184	Sequence 184, App
40	103.5	5.3	730	4	US-09-398-865A-2	Sequence 2, Appli
41	103	5.2	1248	2	US-09-080-897-2	Sequence 2, Appli
42	103	5.2	1248	4	US-09-323-735-2	Sequence 2, Appli
43	103	5.2	1255	2	US-09-080-897-4	Sequence 4, Appli
44	103	5.2	1255	3	US-08-899-595-1	Sequence 1, Appli
45	103	5.2	1255	4	US-09-323-735-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-08-290-919-3
; Sequence 3, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Modified-site
```

```
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-3


  Query Match            13.8%;  Score 272;  DB 1;  Length 53;
  Best Local Similarity   92.3%;  Pred. No. 2.9e-15;
  Matches   48;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    306 PNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSN 357
          ||||||||||||||||||||||||||||||:|||||||||   |||||||
Db      2 PNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSN 53


RESULT   2
US-08-290-919-4
; Sequence 4, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-4


  Query Match            12.9%;  Score 253;  DB 1;  Length 53;
  Best Local Similarity   86.5%;  Pred. No. 9.3e-14;
  Matches   45;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    306 PNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSN 357
          ||||||||||||||| ||||||||: |||||:|||||||||   |||||||
Db      2 PNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDSYPLFDGIFCSSSN 53


RESULT   3
US-08-290-919-11
; Sequence 11, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  106 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-290-919-11


  Query Match            11.8%;  Score 232;  DB 1;  Length 106;
  Best Local Similarity   48.0%;  Pred. No. 1.1e-11;
  Matches   47;  Conservative   11;  Mismatches   36;  Indels    4;  Gaps    2;

Qy    261 KHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQ-EGSKCVENPNPTCNENNGGCDA 319
          ||     |    | :|:|||| :  || :||| ||: ||: |||| ||||: ||||||
Db     12 KHVCVDTRD---IPKNAGCFRDDNGTEEWRCLLGYKKGEGNTCVENNNPTCDINNGGCDP 68

Qy    320 DAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSN 357
          | |   :|  | ||| | | :|       :||||||:
Db     69 TASCQNAESTENSKKIICTCKEPTPNAYYEGVFCSSSS 106


RESULT   4
US-08-290-919-1
```

```
; Sequence 1, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  48 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-1


  Query Match             10.2%;  Score 200;  DB 1;  Length 48;
  Best Local Similarity   87.5%;  Pred. No. 1.4e-09;
  Matches   35;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    270 KTMIQQNSGCFRHLDEREECKCLLNYKQEGSKCVENPNPT 309
          |    ||||||||||||||||||||||||| |||||||||
Db      9 KKQCPQNSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 48


RESULT   5
US-08-290-919-12
; Sequence 12, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  48 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-290-919-12


  Query Match             10.2%;  Score 200;  DB 1;  Length 48;
  Best Local Similarity   87.5%;  Pred. No. 1.4e-09;
  Matches   35;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    270 KTMIQQNSGCFRHLDEREECKCLLNYKQEGSKCVENPNPT 309
          |    ||||||||||||||||||||||||| |||||||||
Db      9 KKQCPQNSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 48


RESULT   6
US-08-290-919-2
; Sequence 2, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/290,919
;       FILING DATE:  04-OCT-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9203821.5
;       FILING DATE:  22-FEB-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/GB93/00367
;       FILING DATE:  22-FEB-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KOKULIS, PAUL N.
;       REGISTRATION NUMBER:  16,773
;       REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 861-3000
;       TELEFAX:  (202) 822-0944
;       TELEX:  6714627 CUSH
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  48 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  1
;       OTHER INFORMATION:  /label= X
;       OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-2


  Query Match             10.0%;  Score 197;  DB 1;  Length 48;
  Best Local Similarity   85.0%;  Pred. No. 2.4e-09;
  Matches   34;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy     270 KTMIQQNSGCFRHLDEREECKCLLNYKQEGSKCVENPNPT 309
           |    :||||||||||||||||||||||||| |||||||||
Db       9 KKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 48


RESULT   7
US-08-257-073-3
; Sequence 3, Application US/08257073
; Patent No. 5766597
;  GENERAL INFORMATION:
;     APPLICANT:  Paoletti, Enzo
;     APPLICANT:  de Taisne, Charles
;     APPLICANT:  Tine, John A.
;     TITLE OF INVENTION:  MALARIA RECOMBINANT POXVIRUS VACCINE
;     NUMBER OF SEQUENCES:  143
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Curtis, Morris & Safford, P.C.
;       STREET:  530 Fifth Avenue, 25th Floor
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  UNITED STATES OF AMERICA
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/257,073
;       FILING DATE:  09-JUN-1994
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/075,783
;       FILING DATE:  11-JUN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/852,305
;       FILING DATE:  18-MAR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/672,183
;       FILING DATE:  20-MAR-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Frommer, William S.
;       REGISTRATION NUMBER:  25,506
;       REFERENCE/DOCKET NUMBER:  454310-2570
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 840-3333
;       TELEFAX:  (212) 840-0712
;       TELEX:  425066 CURTMS
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  984 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FRAGMENT TYPE:  internal
US-08-257-073-3


  Query Match             6.5%;  Score 127;  DB 1;  Length 984;
  Best Local Similarity   20.4%;  Pred. No. 0.051;
  Matches   94;  Conservative   64;  Mismatches  161;  Indels  142;  Gaps   21;

Qy       4 PSVIHNILSKIENEYEVLYLKPLA-----GVYRSLKKQLENN---VMTFNVNVKDILNSP 55
           |  : ||    :|   |:|:|          ||   |   |||   |  : :| |:   :|
Db     236 PRNLQNICETGKNFKLVVYIKENTLIIKWKVYGETKDTTENNKVDVRKYLINEKE---TP 292

Qy      56 FN-------KRENFKNVLES------DLIPYK-DLTSSNYVVKDPYKF-------LNKEK 94
           |       |  |  |::||        || | |  :||  :   :          |  ||
Db     293 FTSILIHAYKEHNGTNLIESKNYALGSDIPEKCDTLASNCFLSGNFNIEKCFQCALLVEK 352

Qy      95 RDKFLSSYNYIKDSIDTDINFANDVLGYYKILSEKYKSDLDSIKKYINDKQGENEKYLPF 154
            :|                    |||   || |||   |    ||    |   ::      .
Db     353 ENK-------------------NDVC--YKYLSEDIVSKFKEIKAETEDDDEDDYTEYKL 391

Qy     155 LNNIET----LYKT--VNDKIDLFVIH-----LEAKVLNYTYEKSNVEIKELIYLKTIQD 203
           :|:     ::||   ||| :|  :      |: :::||     :|:    :    : :
Db     392 TESIDNILVKMFKTNENNDKSELIKLEEVDDSLKLELMNYCSLLKDVDTTGTLDNYGMGN 451

Qy     204 KLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGMVFENL------LKSILSNLLDWKLA 257
           ::  |   ||   :    :: | | |  || :  :  :|:       : ::   |:: |
Db     452 EMDIF---NNLKRLLIYHSEENINTLKNKFRNAAVCLKNVDDWIVNKRGLVLPELNYDLE 508

Qy     258 RYVKHF--------------------TTPMRKKTMIQQNS----------------GC 279
           : :|                      ||  :: |:  ||                  |
Db     509 YFNEHLYNDKNSPEDKDNKGKGVVHVDTTLEKEDTLSYDNSDNMFCNKEYCNRLKDENNC 568

Qy     280 FRHLDEREECKC------LLNYKQEGSKCVENPNPT---------C--NENNGGCDADAK 322
           :|   ::  |          |  |  :|::   ||          |   |:   ||  :
Db     569 ISNLQVEDQGNCDTSWIFASKYHLETIRCMKGYEPTKISALYVANCYKGEHKDRCDEGSS 628

Qy     323 CTE-----EDSG---------SNGKKITCQCTKPDSYPLSM 349
           |      || |          |  |:  || | : : :::
Db     629 PMEFLQIIEDYGFLPAESNYPYNYVKVGEQCPKVEDHWMNL 669


RESULT   8
US-08-184-009-120
; Sequence 120, Application US/08184009
; Patent No. 5833975
;  GENERAL INFORMATION:
;     APPLICANT:  Paoletti, Enzo
;     APPLICANT:  Tartaglia, James
;     APPLICANT:  Cox, William I.
;     TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;     NUMBER OF SEQUENCES:  217
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Curtis, Morris & Safford
;       STREET:  530 Fifth Avenue
```

```
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/184,009
;      FILING DATE:  19-JAN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2530
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  120:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  984 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-184-009-120


  Query Match             6.5%;  Score 127;  DB 2;  Length 984;
  Best Local Similarity   20.4%;  Pred. No. 0.051;
  Matches   94;  Conservative   64;  Mismatches  161;  Indels  142;  Gaps   21;

Qy       4 PSVIHNILSKIENEYEVLYLKPLA-----GVYRSLKKQLENN---VMTFNVNVKDILNSP 55
           |  : ||    :|   |:|:|         ||   |   |||   |  : :| |:   :|
Db     236 PRNLQNICETGKNFKLVVYIKENTLIIKWKVYGETKDTTENNKVDVRKYLINEKE---TP 292

Qy      56 FN-------KRENFKNVLES------DLIPYK-DLTSSNYVVKDPYKF-------LNKEK 94
           |        |  |  |::||        || | |  :||  :   :          |  ||
Db     293 FTSILIHAYKEHNGTNLIESKNYALGSDIPEKCDTLASNCFLSGNFNIEKCFQCALLVEK 352

Qy      95 RDKFLSSYNYIKDSIDTDINFANDVLGYYKILSEKYKSDLDSIKKYINDKQGENEKYLPF 154
            :|                 |||   || |||   |    ||    |   ::
Db     353 ENK------------------NDVC--YKYLSEDIVSKFKEIKAETEDDDEDDYTEYKL 391

Qy     155 LNNIET----LYKT--VNDKIDLFVIH-----LEAKVLNYTYEKSNVEIKELIYLKTIQD 203
             :|:     ::||   ||| :|  :       |: :::||     :|:    :    : :
Db     392 TESIDNILVKMFKTNENNDKSELIKLEEVDDSLKLELMNYCSLLKDVDTTGTLDNYGMGN 451

Qy     204 KLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGMVFENL------LKSILSNLLDWKLA 257
           ::  |   ||   :    :: | | |  || :  :  :|:       : ::   |:: |
Db     452 EMDIF---NNLKRLLIYHSEENINTLKNKFRNAAVCLKNVDDWIVNKRGLVLPELNYDLE 508

Qy     258 RYVKHF--------------------TTPMRKKTMIQQNS----------------GC 279
           : :|                     ||  :: |:   ||                  |
Db     509 YFNEHLYNDKNSPEDKDNKGKGVVHVDTTLEKEDTLSYDNSDNMFCNKEYCNRLKDENNC 568

Qy     280 FRHLDEREECKC------LLNYKQEGSKCVENPNPT---------C--NENNGGCDADAK 322
             :|  ::  |          |  :|::   ||          |   |:   ||  :
Db     569 ISNLQVEDQGNCDTSWIFASKYHLETIRCMKGYEPTKISALYVANCYKGEHKDRCDEGSS 628

Qy     323 CTE-----EDSG---------SNGKKITCQCTKPDSYPLSM 349
             |     || |          |  |:  || | : : :::
Db     629 PMEFLQIIEDYGFLPAESNYPYNYVKVGEQCPKVEDHWMNL 669


RESULT   9
US-08-458-356-120
; Sequence 120, Application US/08458356
; Patent No. 5942235
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;    NUMBER OF SEQUENCES:  217
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,356
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/184,009
;      FILING DATE:  19-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2530
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  120:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  984 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-458-356-120


  Query Match             6.5%;  Score 127;  DB 2;  Length 984;
  Best Local Similarity   20.4%;  Pred. No. 0.051;
  Matches   94;  Conservative   64;  Mismatches  161;  Indels  142;  Gaps   21;

Qy       4 PSVIHNILSKIENEYEVLYLKPLA-----GVYRSLKKQLENN---VMTFNVNVKDILNSP 55
           |  : ||    :|   |:|:|         ||   |   |||   |  : :| |:   :|
Db     236 PRNLQNICETGKNFKLVVYIKENTLIIKWKVYGETKDTTENNKVDVRKYLINEKE---TP 292

Qy      56 FN-------KRENFKNVLES------DLIPYK-DLTSSNYVVKDPYKF-------LNKEK 94
           |        |  |  |::||        || | |  :||  :   :          |  ||
Db     293 FTSILIHAYKEHNGTNLIESKNYALGSDIPEKCDTLASNCFLSGNFNIEKCFQCALLVEK 352

Qy      95 RDKFLSSYNYIKDSIDTDINFANDVLGYYKILSEKYKSDLDSIKKYINDKQGENEKYLPF 154
            :|                 |||   || |||   |    ||    |   ::
Db     353 ENK------------------NDVC--YKYLSEDIVSKFKEIKAETEDDDEDDYTEYKL 391

Qy     155 LNNIET----LYKT--VNDKIDLFVIH-----LEAKVLNYTYEKSNVEIKELIYLKTIQD 203
             :|:     ::||   ||| :|  :       |: :::||     :|:    :    : :
Db     392 TESIDNILVKMFKTNENNDKSELIKLEEVDDSLKLELMNYCSLLKDVDTTGTLDNYGMGN 451

Qy     204 KLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGMVFENL------LKSILSNLLDWKLA 257
           ::  |   ||   :    :: | | |  || :  :  :|:       : ::   |:: |
Db     452 EMDIF---NNLKRLLIYHSEENINTLKNKFRNAAVCLKNVDDWIVNKRGLVLPELNYDLE 508

Qy     258 RYVKHF--------------------TTPMRKKTMIQQNS----------------GC 279
           : :|                     ||  :: |:   ||                  |
Db     509 YFNEHLYNDKNSPEDKDNKGKGVVHVDTTLEKEDTLSYDNSDNMFCNKEYCNRLKDENNC 568

Qy     280 FRHLDEREECKC------LLNYKQEGSKCVENPNPT---------C--NENNGGCDADAK 322
```

```
           :|   ::  |         |  |  :|::   ||         |   |:   ||  :
Db    569 ISNLQVEDQGNCDTSWIFASKYHLETIRCMKGYEPTKISALYVANCYKGEHKDRCDEGSS 628

Qy    323 CTE-----EDSG---------SNGKKITCQCTKPDSYPLSM 349
            |     || |          |  |:  || | : : :::
Db    629 PMEFLQIIEDYGFLPAESNYPYNYVKVGEQCPKVEDHWMNL 669


RESULT 10
US-08-446-855A-2
; Sequence 2, Application US/08446855A
; Patent No. 5849573
;  GENERAL INFORMATION:
;    APPLICANT:  Stewart, Thomas S
;    APPLICANT:  Flores, Maria V
;    APPLICANT:  O'Sullivan, William J
;    TITLE OF INVENTION:  Nucleotide sequence encoding carbamoyl
;    TITLE OF INVENTION:  phosphate synthetase II
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye PC
;      STREET:  1100 No. 5849573th Glebe Road, 8th Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,855A
;      FILING DATE:  06-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mitchard, Leonard C
;      REGISTRATION NUMBER:  29.009
;      REFERENCE/DOCKET NUMBER:  47-80
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4000
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2391 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-446-855A-2


  Query Match             6.0%;  Score 117.5;  DB 2;  Length 2391;
  Best Local Similarity   20.1%;  Pred. No. 0.95;
  Matches   69;  Conservative   60;  Mismatches  111;  Indels  103;  Gaps   16;

Qy      37 LENNV---MTFNVNVKDILNSPFNKRENFKNVLESDLIPYKDLTSSNYVV--------KD 85
           :|||:   :: | |:|  :|:  |   | :|| |::|    :   |:::          |
Db    1333 IENNICHDISINKNIKVTINNSNNSISNNENV-ETNLNCVSERAGSHHIYGKEEKSIGSD 1391

Qy      86 PYKFLNKEKRDKFLSSYNYIKDSIDTDIN-FANDVLGYYKILSEKYKSDLDSIKKY---- 140
             |: :  :   |  |   :  : |:|   ||          | : |:::   :
Db    1392 DTNILSAQNSNNNFSCNNENMNKANVDVNVLENDT---------KKREDINTTTVFMEGQ 1442

Qy     141 ---INDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKVLNYTYEKSNVEI-KELI 196
              ||:|  ||   |               |: |: :::|: |  ||    :||:  |: :
Db    1443 NSVINNKNKENSSLLK------------GDEEDIVMVNLK-KENNYNSVINNVDCRKKDM 1489

Qy     197 YLKTIQDKLADFKK--------NNNFVGIADLSTDYNHNNLLTKFLSTGMVFENLLKSIL 248
           | | |:   :||         ||| |      | :: |: |
Db    1490 DGKNINDECKTYKKNKYKDMGLNNNIVDELSNGTSHSTNDHL------------------ 1531

Qy     249 SNLLDWKLARYVKHFTTP--------------MRKKTMIQQNSG-CFRHLDEREECKCLL 293
```

```
                |: :| |                 ::|::  :| |  :  :|     :  :
Db   1532 ----------YLDNFNTSDEEIGNNKNMDMYLSKEKSISNKNPGNSYYVVDSVYNNEYKI 1581

Qy    294 NYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGS---NGK 333
          |  :|    ::| |     ||   : :  |:  :: |   |||
Db   1582 NKMKE---LIDNENLNDEYNN---NVNMNCSNYNNASAFVNGK 1618


RESULT 11
US-09-150-741-2
; Sequence 2, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
;  APPLICANT: Stewart et al.
;  TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
;  TITLE OF INVENTION: Synthetase II
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/150,741
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PL6380
;  EARLIER FILING DATE: 1992-12-16
;  EARLIER APPLICATION NUMBER: AU93/00617
;  EARLIER FILING DATE: 1993-12-02
;  EARLIER APPLICATION NUMBER: 08/446,855
;  EARLIER FILING DATE: 1995-07-06
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2391
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2


  Query Match             6.0%;  Score 117.5;  DB 4;  Length 2391;
  Best Local Similarity   20.1%;  Pred. No. 0.95;
  Matches   69;  Conservative   60;  Mismatches  111;  Indels  103;  Gaps   16;

Qy      37 LENNV---MTFNVNVKDILNSPFNKRENFKNVLESDLIPYKDLTSSNYVV--------KD 85
           :|||:   :: | |:|  :|:  |   | :|| |::|    :   |:::          |
Db    1333 IENNICHDISINKNIKVTINNSNNSISNNENV-ETNLNCVSERAGSHHIYGKEEKSIGSD 1391

Qy      86 PYKFLNKEKRDKFLSSYNYIKDSIDTDIN-FANDVLGYYKILSEKYKSDLDSIKKY---- 140
             |: :  :   |  |   :  : |:|   ||          | : |:::   :
Db    1392 DTNILSAQNSNNNFSCNNENMNKANVDVNVLENDT---------KKREDINTTTVFMEGQ 1442

Qy     141 ---INDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKVLNYTYEKSNVEI-KELI 196
              ||:|  ||   |               |: |: :::|: |  ||    :||:  |: :
Db    1443 NSVINNKNKENSSLLK------------GDEEDIVMVNLK-KENNYNSVINNVDCRKKDM 1489

Qy     197 YLKTIQDKLADFKK--------NNNFVGIADLSTDYNHNNLLTKFLSTGMVFENLLKSIL 248
           | | |:   :||         ||| |      | :: |: |
Db    1490 DGKNINDECKTYKKNKYKDMGLNNNIVDELSNGTSHSTNDHL------------------ 1531

Qy     249 SNLLDWKLARYVKHFTTP--------------MRKKTMIQQNSG-CFRHLDEREECKCLL 293
                |: :| |                 ::|::  :| |  :  :|     :  :
Db    1532 ----------YLDNFNTSDEEIGNNKNMDMYLSKEKSISNKNPGNSYYVVDSVYNNEYKI 1581

Qy     294 NYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGS---NGK 333
           |  :|    ::| |     ||   : :  |:  :: |   |||
Db    1582 NKMKE---LIDNENLNDEYNN---NVNMNCSNYNNASAFVNGK 1618


RESULT 12
US-08-323-170B-2
; Sequence 2, Application US/08323170B
; Patent No. 5733772
;  GENERAL INFORMATION:
;    APPLICANT:  Williamson, Kim C.
;    APPLICANT:  Kaslow, David C.
```

```
;    TITLE OF INVENTION:  Cloning and Expression of Plasmodium
;    TITLE OF INVENTION:  faliciparum Transmission-Blocking Target Antigen, Pfs230
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/323,170B
;      FILING DATE:  13-OCT-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/010,409
;      FILING DATE:  29-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Quine, Jonathan A.
;      REGISTRATION NUMBER:  P-41,261
;      REFERENCE/DOCKET NUMBER:  015280-113100US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3135 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-323-170B-2


  Query Match             5.9%;  Score 117;  DB 1;  Length 3135;
  Best Local Similarity   22.2%;  Pred. No. 1.5;
  Matches   64;  Conservative   41;  Mismatches   83;  Indels  100;  Gaps   13;

Qy      21 LYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRENFKNVLESDLIPYKDLTSSN 80
           |:|  |  :: | || |  ||:  | |:||  :  |:            |  |     |:
Db      10 LFLIILVYIF-SEKKDLRCNVIKGN-NIKDDEDKRFH-----------LFYY-----SH 50

Qy      81 YVVKDPYKFLNKEKRDKFLSS---YNYIKD---SIDTDINF-----------------AN 117
             : | |     | |:: |  :   ||  |:     || :                   :
Db      51 NLFKTPETKEKKNKKECFYKNGGIYNLSKEIRMRKDTSVKIKQRTCPFHKEGSSFEMGSK 110

Qy     118 DVLGYYKILSEKYKSDLDSI---KKYIND-------KQGENEKYLPFLNNIETLYKTVND 167
           ::  :| |: :| :  ||:|   |   ||           || :  : ||:   |  ||
Db     111 NITCFYPIVGKKERKTLDTIIIKKNVTNDHVVSSDMHSNVQEKNMILIRNIDKENK--ND 168

Qy     168 KIDLFVIHLEAKVLNYTYEKSNVEIKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHN 227
                : ::| |:   |||                                 : ||  :
Db     169 -----IQNVEEKIQRDTYE---------------------------------NKDYESD 189

Qy     228 NLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQ 275
           : | ::       || | : |   |       :| |::| ||||::|:
Db     190 DTLIEWFDDNTNEENFLLTFLKRCL-------MKIFSSPKRKKTVVQK 230


RESULT  13
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT: DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 0660-0125-0 PCT
;  CURRENT APPLICATION NUMBER: US/08/973,462B
;  CURRENT FILING DATE: 1998-02-06
;  EARLIER APPLICATION NUMBER: PCT/FR96/00894
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: FR 95/07007
;  EARLIER FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 1786
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Polypeptide
US-08-973-462-8


  Query Match             5.8%;  Score 114;  DB 4;  Length 1786;
  Best Local Similarity   25.4%;  Pred. No. 1.2;
  Matches   67;  Conservative   43;  Mismatches   84;  Indels   70;  Gaps   16;

Qy       5 SVIHNILSKIENEYEVLY-----LKPLAGVYRSLKKQLENNVMTFNVNV---KD----IL 52
             :|: :||  |||  | |       :    ||  :: : :| ||   :|:|   ||    ||
Db    1046 NVVSSILDNIENMKEGLLNKLENISSTEGVQETVTEHVEQNVYV-DVDVPAMKDQFLGIL 1104

Qy      53 NSP-------FNKRENFKNVLESDLIPYKDLTSSNYVVKDPYK---------------FL 90
           |          ||  : ||:  |||:|  :::     | |:  |               |
Db    1105 NEAGGLKEMFFNLEDVFKS--ESDVITVEEI-KDEPVQKEVEKETVSIIEEMEENIVDVL 1161

Qy      91 NKEKR---DKFLSSYNYIKDSIDTDINFANDVLGYYKILSEKYKSDLDSIKKYI--NDKQ 145
              :||    || :       |:::  |  ::|     : : :| | |:  : : :  ||
Db    1162 EEEKEDLTDKMI-------DAVEESIEISSDSKEETESIKDKEK-DVSLVVEEVQDNDMD 1213

Qy     146 GENEKYLPFLNNIETLYK---TVNDKIDLFVIHLEAKVLNYTYEKSNVE---IKELIYLK 199
             || |   |  | | |    :||        : :|::  | | : ||   ||::  ||
Db    1214 ESVEKVLELKNMEEELMKDAVEIND--------ITSKLIEETQELNEVEADLIKDMEKLK 1265

Qy     200 TIQDKLADFKKNNNFVGIADLSTD 223
           ::  |::  |      | |   |
Db    1266 ELEKALSEDSKE-----IIDAKDD 1284


RESULT  14
US-08-392-625-21
; Sequence 21, Application US/08392625
; Patent No. 5837485
;  GENERAL INFORMATION:
;    APPLICANT:  Entian, Karl-Dieter
;    APPLICANT:  G tz, Friedrich
;    APPLICANT:  Schnell, No. 5837485bert
;    APPLICANT:  Augustin, Johannes
;    APPLICANT:  Engelke, Germar
;    APPLICANT:  Rosenstein, Ralf
;    APPLICANT:  Kaletta, Cortina
;    APPLICANT:  Klein, Cora
;    APPLICANT:  Wieland, Bernd
;    APPLICANT:  Kupke, Thomas
;    APPLICANT:  Jung, G nther
;    APPLICANT:  Kellner, Roland
;    TITLE OF INVENTION:  Biosynthetic Process For The Preparation
;    TITLE OF INVENTION:  Of Chemical Compounds
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/392,625
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/876,791
;      FILING DATE:  30-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0652.0980002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  455 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-392-625-21


  Query Match             5.7%;  Score 113;  DB 2;  Length 455;
  Best Local Similarity   22.2%;  Pred. No. 0.24;
  Matches   80;  Conservative   51;  Mismatches  119;  Indels  110;  Gaps   19;

Qy      17 EYEVLYLKPLAGVYRSLKK--QLEN-NVMTFNVNVKDILNSPFNKRENFKNVLESDLIPY 73
           || || ||     |  | |  :|:|  ::  |:|             | | :||: :
Db      11 EYSVLILKKKNLFYLFLMKLQKLKNIGMVVININ------------NIKKILENKITFL 57

Qy      74 KDLTSSNYVVK------DPYK----------FLNKEKR------DKFLSSY-----NYIK 106
            |:  : |:::      |||           ||:  ::      :| :  |      |::
Db      58 SDIEKATYIIENQSEYWDPYTLSHGYPGIILFLSASEKVFHKDLEKVIHQYIRKLGPYLE 117

Qy     107 DSID--------TDINFANDVLGYYKILSEK---YKSDLDSIK----KYINDKQGENEKY 151
             ||        : | || |      | |:|   |:| |: |     :|: |
Db     118 SGIDGFSLFSGLSGIGFALD------IASDKQYSYQSILEQIDNLLVQYVFD-------- 163

Qy     152 LPFLNN-----IETLYKTVNDKIDLFVIHLEAKVLNYTYEKSNVEIKELIYLKTIQDKLA 206
             ||||       | |  :      :    |      ||  :|:   |  | | |||
Db     164 --FLNNDALEVTPTNYDIIQGFSGVGRYLLNRISYNYAKKALKHI--LNYFKTIH---- 215

Qy     207 DFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM---VFENLLKSILSNLLDWKL---ARYV 260
           : |:|  |       | :  |  : :: |:   :   |  : || :   ::     ::
Db     216 -YSKDNWLVSNEHQFLDIDKQNFPSGNINLGLAHGILGPLSLTALSKMNGIEIEGHEEFL 274

Qy     261 KHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQEGSKCVENPNPTCNENNGGCDAD 320
           : ||: : |    : |:  |   |                  :||  |  :  || |  |
Db     275 QDFTSFLLKPE-FKNNNEWFDRYD----------------ILENYIPNYSVRNGWCYGD 316


RESULT  15
US-08-466-961A-21
; Sequence 21, Application US/08466961A
; Patent No. 5843709
;  GENERAL INFORMATION:
;    APPLICANT:  Entian, Karl-Dieter
;    APPLICANT:  G tz, Friedrich
;    APPLICANT:  Schnell, No. 5843709bert
;    APPLICANT:  Augustin, Johannes
;    APPLICANT:  Engelke, Germar
;    APPLICANT:  Rosenstein, Ralf
;    APPLICANT:  Kaletta, Cortina
;    APPLICANT:  Klein, Cora
;    APPLICANT:  Wieland, Bernd
;    APPLICANT:  Kupke, Thomas
;    APPLICANT:  Jung, G nther
;    APPLICANT:  Kellner, Roland
;    TITLE OF INVENTION:  Biosynthetic Process for the Preparation of
;    TITLE OF INVENTION:  Chemical Compounds
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, NW
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,961A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/392,625
;      FILING DATE:  22-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/876,791
;      FILING DATE:  30-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/784,234
;      FILING DATE:  31-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0652.0980004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  455 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-466-961A-21


  Query Match             5.7%;  Score 113;  DB 2;  Length 455;
  Best Local Similarity   22.2%;  Pred. No. 0.24;
  Matches   80;  Conservative   51;  Mismatches  119;  Indels  110;  Gaps   19;

Qy      17 EYEVLYLKPLAGVYRSLKK--QLEN-NVMTFNVNVKDILNSPFNKRENFKNVLESDLIPY 73
           || || ||     |  | |  :|:|  ::  |:|             | | :||: :
Db      11 EYSVLILKKKNLFYLFLMKLQKLKNIGMVVININ------------NIKKILENKITFL 57

Qy      74 KDLTSSNYVVK------DPYK----------FLNKEKR------DKFLSSY-----NYIK 106
            |:  : |:::      |||           ||:  ::      :| :  |      |::
Db      58 SDIEKATYIIENQSEYWDPYTLSHGYPGIILFLSASEKVFHKDLEKVIHQYIRKLGPYLE 117

Qy     107 DSID--------TDINFANDVLGYYKILSEK---YKSDLDSIK----KYINDKQGENEKY 151
             ||        : | || |      | |:|   |:| |: |     :|: |
Db     118 SGIDGFSLFSGLSGIGFALD------IASDKQYSYQSILEQIDNLLVQYVFD-------- 163

Qy     152 LPFLNN-----IETLYKTVNDKIDLFVIHLEAKVLNYTYEKSNVEIKELIYLKTIQDKLA 206
             ||||       | |  :      :    |      ||  :|:   |  | | |||
Db     164 --FLNNDALEVTPTNYDIIQGFSGVGRYLLNRISYNYAKKALKHI--LNYFKTIH---- 215

Qy     207 DFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM---VFENLLKSILSNLLDWKL---ARYV 260
           : |:|  |       | :  |  : :: |:   :   |  : || :   ::     ::
Db     216 -YSKDNWLVSNEHQFLDIDKQNFPSGNINLGLAHGILGPLSLTALSKMNGIEIEGHEEFL 274

Qy     261 KHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQEGSKCVENPNPTCNENNGGCDAD 320
           : ||: : |    : |:  |   |                  :||  |  :  || |  |
Db     275 QDFTSFLLKPE-FKNNNEWFDRYD----------------ILENYIPNYSVRNGWCYGD 316


Search completed: August  8, 2001, 12:30:01
Job time: 49 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:33:42 ; Search time 70.07 Seconds
(without alignments)
743.944 Million cell updates/sec

Title: US-09-500-376-3
Perfect score: 2052
Sequence: 1 AISVTMDNILSGFENEYDVI.....SNFLGISFLLILMLILYSFI 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2033	99.1	570	5	Q25968	Q25968 plasmodium
2	2033	99.1	570	5	Q9TYG2	Q9tyg2 plasmodium
3	2017.5	98.3	569	5	Q25970	Q25970 plasmodium
4	2017.5	98.3	569	5	Q25980	Q25980 plasmodium
5	2017.5	98.3	569	5	Q25982	Q25982 plasmodium
6	2015	98.2	1720	5	Q25922	Q25922 plasmodium
7	2013.5	98.1	569	5	Q25978	Q25978 plasmodium
8	2011	98.0	652	5	Q25923	Q25923 plasmodium
9	2001.5	97.5	569	5	Q25969	Q25969 plasmodium
10	2001.5	97.5	569	5	Q25974	Q25974 plasmodium
11	2001.5	97.5	569	5	Q25975	Q25975 plasmodium
12	2001.5	97.5	569	5	Q25977	Q25977 plasmodium
13	2001.5	97.5	569	5	Q25979	Q25979 plasmodium
14	2001.5	97.5	1694	5	Q9TZT5	Q9tzt5 plasmodium
15	2001.5	97.5	1694	5	Q9NHX1	Q9nhx1 plasmodium
16	2001.5	97.5	1704	5	Q9TZT4	Q9tzt4 plasmodium
17	1997.5	97.3	569	5	Q25983	Q25983 plasmodium
18	1987.5	96.9	651	5	Q25924	Q25924 plasmodium
19	1982.5	96.6	569	5	Q25967	Q25967 plasmodium
20	1930	94.1	373	5	Q25727	Q25727 plasmodium
21	1925	93.8	373	5	Q25728	Q25728 plasmodium
22	1919.5	93.5	372	5	Q25717	Q25717 plasmodium
23	1914.5	93.3	372	5	Q25718	Q25718 plasmodium
24	1914.5	93.3	372	5	Q25719	Q25719 plasmodium
25	1914.5	93.3	372	5	Q25720	Q25720 plasmodium
26	1910	93.1	373	5	O43996	O43996 plasmodium
27	1905	92.8	373	5	Q25724	Q25724 plasmodium
28	1900	92.6	373	5	O43995	O43995 plasmodium
29	1900	92.6	373	5	Q25721	Q25721 plasmodium
30	1898	92.5	373	5	Q25723	Q25723 plasmodium
31	1891	92.2	373	5	Q25722	Q25722 plasmodium
32	1887.5	92.0	372	5	Q25725	Q25725 plasmodium
33	1882.5	91.7	372	5	Q25726	Q25726 plasmodium
34	1881.5	91.7	372	5	O43997	O43997 plasmodium
35	1257.5	61.3	539	5	Q25972	Q25972 plasmodium
36	1257.5	61.3	539	5	Q25981	Q25981 plasmodium
37	1254.5	61.1	400	5	Q03999	Q03999 plasmodium
38	1254.5	61.1	539	5	Q25966	Q25966 plasmodium
39	1254.5	61.1	539	5	Q25976	Q25976 plasmodium
40	1254.5	61.1	539	5	Q25984	Q25984 plasmodium
41	1254.5	61.1	539	5	Q9TYG1	Q9tyg1 plasmodium
42	1248.5	60.8	539	5	Q25971	Q25971 plasmodium
43	1248.5	60.8	539	5	Q25973	Q25973 plasmodium
44	1142	55.7	218	5	Q9TVG8	Q9tvg8 plasmodium
45	828	40.4	1726	5	Q02569	Q02569 plasmodium

ALIGNMENTS

RESULT 1
Q25968
ID Q25968 PRELIMINARY; PRT; 570 AA.
AC Q25968;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13345; BAA02606.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1 1
SQ SEQUENCE 570 AA; 64632 MW; 424BF553CCC2F2BE CRC64;

Query Match 99.1%; Score 2033; DB 5; Length 570;
Best Local Similarity 99.2%; Pred. No. 2.3e-101;
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    297 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 356

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    357 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 416

Qy    241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    417 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 476

Qy    301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    477 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 536

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db    537 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 570


RESULT   2
Q9TYG2
ID   Q9TYG2      PRELIMINARY;      PRT;   570 AA.
AC   Q9TYG2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88011243; PubMed=3079521;
RA   Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT   "Allelic dimorphism in a surface antigen gene of the malaria parasite
RT   Plasmodium falciparum.";
RL   J. Mol. Biol. 195:273-287(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13343; BAA02604.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   570 AA;  64630 MW;  8674DEC89B2D662A CRC64;


  Query Match             99.1%;  Score 2033;  DB 5;  Length 570;
  Best Local Similarity   99.2%;  Pred. No. 2.3e-101;
  Matches  391;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    297 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 356

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    357 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 416

Qy    241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    417 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 476

Qy    301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    477 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 536

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db    537 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 570


RESULT   3
Q25970
ID   Q25970         PRELIMINARY;   PRT;   569 AA.
AC   Q25970;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13347; BAA02608.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64502 MW;  4D761FF472753142 CRC64;


  Query Match             98.3%;  Score 2017.5;  DB 5;  Length 569;
  Best Local Similarity   99.0%;  Pred. No. 1.5e-100;
  Matches  390;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 179
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db    297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy    180 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy    240 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 299
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy    300 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 359
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 535

Qy    360 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db    536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569
```

```
RESULT   4
Q25980
ID   Q25980         PRELIMINARY;   PRT;   569 AA.
AC   Q25980;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13353; BAA02614.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64536 MW;  8008861DECECD8DC CRC64;


  Query Match             98.3%;  Score 2017.5;  DB 5;  Length 569;
  Best Local Similarity   99.0%;  Pred. No. 1.5e-100;
  Matches  390;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db     297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 535

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db     536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT   5
Q25982
ID   Q25982         PRELIMINARY;   PRT;   569 AA.
AC   Q25982;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13354; BAA02615.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64530 MW;  4D734580516D2142 CRC64;


  Query Match             98.3%;  Score 2017.5;  DB 5;  Length 569;
  Best Local Similarity   99.0%;  Pred. No. 1.5e-100;
  Matches  390;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db     297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 535

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db     536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT   6
Q25922
ID   Q25922       PRELIMINARY;       PRT;  1720 AA.
AC   Q25922;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PRECURSOR OF THE MAJOR MEROZOITE SURFACE ANTIGENS.
OS   Plasmodium falciparum (isolate NF54).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5843;
RN   [1]
RP   SEQUENCE OF 1069-1720 FROM N.A.
RC   STRAIN=NF54;
RA   Tolle R., Bujard H., Cooper J.A.;
RL   Exp. Parasitol. 0:0-0(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NF54;
RA   Tolle R.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=NF54;
RX   MEDLINE=96123395; PubMed=8577332;
RA   Pan W., Tolle R., Bujard H.;
RT   "A direct and rapid sequencing strategy for the Plasmodium falciparum
RT   antigen gene gp190/MSA1.";
RL   Mol. Biochem. Parasitol. 73:241-244(1995).
DR   EMBL; Z35327; CAA84556.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Signal; Merozoite.
FT   SIGNAL       1     19       POTENTIAL.
SQ   SEQUENCE   1720 AA;  195725 MW;  717B2FD1E637A8A3 CRC64;


  Query Match             98.2%;  Score 2015;  DB 5;  Length 1720;
  Best Local Similarity   98.5%;  Pred. No. 6.6e-100;
  Matches  388;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    1327 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 1386

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    1387 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1446

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1447 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1506

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           |||||||||||||||||||||||||||||||||||||||||||||||||  |||||||
Db    1507 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFEAIKK 1566

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1567 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1626

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1627 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 1686

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db    1687 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1720


RESULT   7
Q25978
ID   Q25978       PRELIMINARY;   PRT;   569 AA.
AC   Q25978;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13351; BAA02612.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER      1      1
SQ   SEQUENCE   569 AA;  64536 MW;  B8B9B63EECB1DA51 CRC64;


  Query Match             98.1%;  Score 2013.5;  DB 5;  Length 569;
  Best Local Similarity   98.7%;  Pred. No. 2.5e-100;
  Matches  389;  Conservative    0;  Mismatches    4;  Indels    1;  Gaps    1;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 236

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 535

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db     536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT   8
Q25923
ID   Q25923       PRELIMINARY;      PRT;   652 AA.
AC   Q25923;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MEROZOITE SURFACE ANTIGEN 1 (FRAGMENT).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE OF 1-298 FROM N.A.
RC   STRAIN=FCH5/NF7;
RX   MEDLINE=88142999; PubMed=2449612;
RA   Peterson G.M., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
RA   Brown G.V., Anders R.F., Kemp D.J.;
RT   "Variation in the precursor to the major merozoite surface antigens of
RT   Plasmodium falciparum.";
RL   Mol. Biochem. Parasitol. 27:291-302(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FCH5/NF7;
RA   Tolle R.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FCH5/NF7;
RX   MEDLINE=95354793; PubMed=7628566;
RA   Tolle R., Bujard H., Cooper J.A.;
RT   "Plasmodium falciparum: variations within the C-terminal region of
RT   merozoite surface antigen-1.";
RL   Exp. Parasitol. 81:47-54(1995).
DR   EMBL; Z35328; CAA84557.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite.
```

```
FT   NON_TER       1      1
SQ   SEQUENCE   652 AA;  74292 MW;  2B6A87737B490A62 CRC64;


  Query Match             98.0%;  Score 2011;  DB 5;  Length 652;
  Best Local Similarity   98.2%;  Pred. No. 3.9e-100;
  Matches  387;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     259 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 318

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     319 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 378

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     379 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 438

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           |||||||||||||||||||||||||||||||||||||||||||||||||  |||||||
Db     439 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFEAIKK 498

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     499 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 558

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
           |||||||||||||||||| ||||||||||||||||||||:|||||||||||||||||||||
Db     559 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEKDSGSSRKKITCECTK 618

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db     619 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 652


RESULT   9
Q25969
ID   Q25969         PRELIMINARY;   PRT;   569 AA.
AC   Q25969;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13346; BAA02607.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64485 MW;  A26C45914AF01396 CRC64;


  Query Match             97.5%;  Score 2001.5;  DB 5;  Length 569;
  Best Local Similarity   98.2%;  Pred. No. 1.1e-99;
  Matches  387;  Conservative    1;  Mismatches    5;  Indels    1;  Gaps    1;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
           |||||||||||||||||| ||||||||||||||||| |||||||: |||||||||||||||
Db     476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 535

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db     536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT  10
Q25974
ID   Q25974         PRELIMINARY;   PRT;   569 AA.
AC   Q25974;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13348; BAA02609.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64457 MW;  A2691FE569E80396 CRC64;


  Query Match             97.5%;  Score 2001.5;  DB 5;  Length 569;
  Best Local Similarity   98.2%;  Pred. No. 1.1e-99;
  Matches  387;  Conservative    1;  Mismatches    5;  Indels    1;  Gaps    1;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415
```

```
Qy    241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy    301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
          |||||||||||||||||| |||||||||||||||| ||||||||: |||||||||
Db    476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 535

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db    536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT  11
Q25975
ID   Q25975         PRELIMINARY;   PRT;   569 AA.
AC   Q25975;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13349; BAA02610.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64459 MW;  ED4284B2867C9703 CRC64;


  Query Match             97.5%;  Score 2001.5;  DB 5;  Length 569;
  Best Local Similarity   98.2%;  Pred. No. 1.1e-99;
  Matches  387;  Conservative    1;  Mismatches    5;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy    241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy    301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
          |||||||||||||||||| |||||||||||||||| ||||||||: |||||||||
Db    476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 535

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db    536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT  12
Q25977
ID   Q25977         PRELIMINARY;   PRT;   569 AA.
AC   Q25977;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13350; BAA02611.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64485 MW;  A26C45914AF01396 CRC64;


  Query Match             97.5%;  Score 2001.5;  DB 5;  Length 569;
  Best Local Similarity   98.2%;  Pred. No. 1.1e-99;
  Matches  387;  Conservative    1;  Mismatches    5;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy    241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy    301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
          |||||||||||||||||| |||||||||||||||| ||||||||: |||||||||
Db    476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 535

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db    536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT  13
Q25979
ID   Q25979         PRELIMINARY;   PRT;   569 AA.
AC   Q25979;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```

```
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13352; BAA02613.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64457 MW;  A2691FE569E80396 CRC64;


  Query Match             97.5%;  Score 2001.5;  DB 5;  Length 569;
  Best Local Similarity   98.2%;  Pred. No. 1.1e-99;
  Matches  387;  Conservative    1;  Mismatches    5;  Indels    1;  Gaps    1;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSSRKKITCECTK 360
           |||||||||||||||||||| |||||||||||||||||| ||||||||: |||||||||
Db     476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 535

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db     536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT  14
Q9TZT5
ID   Q9TZT5      PRELIMINARY;      PRT;  1694 AA.
AC   Q9TZT5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR.
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HN1;
RA   Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF062348; AAC72884.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite.
SQ   SEQUENCE   1694 AA;  192794 MW;  84CFC0E709F5673B CRC64;


  Query Match             97.5%;  Score 2001.5;  DB 5;  Length 1694;
  Best Local Similarity   98.2%;  Pred. No. 3.4e-99;
  Matches  387;  Conservative    1;  Mismatches    5;  Indels    1;  Gaps    1;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1302 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 1361

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    1362 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1421

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    1422 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1480

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1481 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1540

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    1541 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1600

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSSRKKITCECTK 360
           |||||||||||||||||||| |||||||||||||||||| ||||||||: |||||||||
Db    1601 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1660

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db    1661 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1694


RESULT  15
Q9NHX1
ID   Q9NHX1      PRELIMINARY;      PRT;  1694 AA.
AC   Q9NHX1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MAJOR MEROZOITE SURFACE ANTIGEN.
GN   GP195.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FCC1/HN;
RA   Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;
RT   "Molecular cloning and sequence analysis of major merozoite surface
RT   antigen(gp195)gene of Plasmodium falciparum isolate FCC1/HN.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF218248; AAF27526.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR003247; -.
DR   Pfam; PF00008; EGF; 1.
DR   ProDom; PD001527; -; 1.
KW   Merozoite.
SQ   SEQUENCE   1694 AA;  192766 MW;  B51634A49E0F6728 CRC64;


  Query Match             97.5%;  Score 2001.5;  DB 5;  Length 1694;
  Best Local Similarity   98.2%;  Pred. No. 3.4e-99;
  Matches  387;  Conservative    1;  Mismatches    5;  Indels    1;  Gaps    1;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1302 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 1361

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    1362 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1421
```

```
Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    1422 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1480

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1481 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1540

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1541 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1600

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSSRKKITCECTK 360
           |||||||||||||||||| |||||||||||||||| |||||||: |||||||||
Db    1601 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1660

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db    1661 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1694
```

Search completed: August 8, 2001, 12:33:44
Job time: 272 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:56 ; Search time 34.41 Seconds
(without alignments)
235.763 Million cell updates/sec

Title: US-09-500-376-3
Perfect score: 2052
Sequence: 1 AISVTMDNILSGFENEYDVI.....SNFLGISFLLILMLILYSFI 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	298	14.5	53	1	US-08-290-919-4	Sequence 4, Appli
2	282	13.7	53	1	US-08-290-919-3	Sequence 3, Appli
3	271	13.2	48	1	US-08-290-919-12	Sequence 12, Appl
4	268	13.1	48	1	US-08-290-919-2	Sequence 2, Appli
5	265	12.9	48	1	US-08-290-919-1	Sequence 1, Appli
6	255.5	12.5	106	1	US-08-290-919-11	Sequence 11, Appl
7	127.5	6.2	740	1	US-08-257-073-5	Sequence 5, Appli
8	127.5	6.2	2710	1	US-08-480-604A-6	Sequence 6, Appli
9	127.5	6.2	2710	2	US-08-405-496A-6	Sequence 6, Appli
10	126.5	6.2	798	2	US-07-728-215-30	Sequence 30, Appl
11	120.5	5.9	1494	3	US-08-755-587-186	Sequence 186, App
12	118	5.8	1394	4	US-09-213-053-2	Sequence 2, Appli
13	116	5.7	540	1	US-08-286-325A-2	Sequence 2, Appli
14	115.5	5.6	677	3	US-08-480-640A-115	Sequence 115, App
15	115.5	5.6	677	3	US-08-480-640A-193	Sequence 193, App
16	115.5	5.6	677	3	US-08-295-802-115	Sequence 115, App
17	115.5	5.6	677	4	US-08-686-968C-58	Sequence 58, Appl
18	115.5	5.6	677	4	US-08-686-968C-193	Sequence 193, App
19	115.5	5.6	677	4	US-08-488-237A-115	Sequence 115, App
20	115.5	5.6	677	4	US-08-488-237A-193	Sequence 193, App
21	115.5	5.6	984	1	US-08-257-073-3	Sequence 3, Appli
22	115.5	5.6	984	2	US-08-184-009-120	Sequence 120, App
23	115.5	5.6	984	2	US-08-458-356-120	Sequence 120, App
24	115.5	5.6	1786	4	US-08-973-462-8	Sequence 8, Appli
25	115.5	5.6	2482	1	US-08-328-254-6	Sequence 6, Appli
26	115.5	5.6	3248	1	US-08-353-700-1	Sequence 1, Appli
27	115.5	5.6	3248	5	PCT-US95-16216-1	Sequence 1, Appli
28	112.5	5.5	912	1	US-07-789-915A-8	Sequence 8, Appli
29	112.5	5.5	912	1	US-08-005-002C-8	Sequence 8, Appli
30	112.5	5.5	912	1	US-08-487-203A-8	Sequence 8, Appli
31	112.5	5.5	1657	1	US-08-287-959-1	Sequence 1, Appli
32	111.5	5.4	1093	4	US-09-315-793-52	Sequence 52, Appl
33	111	5.4	1312	2	US-08-592-126-148	Sequence 148, App
34	111	5.4	1312	2	US-08-687-080-51	Sequence 51, Appl
35	110	5.4	976	4	US-09-104-324B-4	Sequence 4, Appli
36	110	5.4	984	1	US-08-242-932-2	Sequence 2, Appli
37	110	5.4	984	1	US-08-714-481-2	Sequence 2, Appli
38	110	5.4	984	5	PCT-US95-06111-2	Sequence 2, Appli
39	109.5	5.3	921	2	US-08-568-459A-8	Sequence 8, Appli
40	109.5	5.3	921	2	US-08-487-826B-8	Sequence 8, Appli
41	109	5.3	1572	2	US-08-290-731C-5	Sequence 5, Appli
42	109	5.3	1579	3	US-08-755-587-184	Sequence 184, App
43	109	5.3	1596	3	US-09-356-952-3	Sequence 3, Appli
44	108.5	5.3	699	3	US-08-851-843A-52	Sequence 52, Appl
45	108.5	5.3	699	4	US-08-974-549A-188	Sequence 188, App

ALIGNMENTS

```
RESULT   1
US-08-290-919-4
; Sequence 4, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Modified-site
```

```
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-4


  Query Match            14.5%;  Score 298;  DB 1;  Length 53;
  Best Local Similarity  100.0%;  Pred. No. 8.3e-16;
  Matches   52;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    325 PNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDSYPLFDGIFCSSSN 376
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 PNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDSYPLFDGIFCSSSN 53


RESULT   2
US-08-290-919-3
; Sequence 3, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-3


  Query Match            13.7%;  Score 282;  DB 1;  Length 53;
  Best Local Similarity  94.2%;  Pred. No. 1.3e-14;
  Matches   49;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    325 PNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDSYPLFDGIFCSSSN 376
          |||||||||||||||| |||||||: |||||||||||||||||||||||||
Db      2 PNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSN 53


RESULT   3
US-08-290-919-12
; Sequence 12, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  48 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-290-919-12


  Query Match            13.2%;  Score 271;  DB 1;  Length 48;
  Best Local Similarity  95.8%;  Pred. No. 7.3e-14;
  Matches   46;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    281 NISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCEENPNPT 328
          |||||||||||||:|||||||||||||||||||||||||| ||||||
Db      1 NISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 48


RESULT   4
US-08-290-919-2
; Sequence 2, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
```

```
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  48 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-2


  Query Match             13.1%;  Score 268;  DB 1;  Length 48;
  Best Local Similarity   97.9%;  Pred. No. 1.2e-13;
  Matches   46;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     282 ISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCEENPNPT 328
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       2 ISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 48


RESULT   5
US-08-290-919-1
; Sequence 1, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  48 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-1


  Query Match             12.9%;  Score 265;  DB 1;  Length 48;
  Best Local Similarity   95.7%;  Pred. No. 2e-13;
  Matches   45;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     282 ISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCEENPNPT 328
           |||||||||||||:||||||||||||||||||||||||||||| ||||||
Db       2 ISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 48


RESULT   6
US-08-290-919-11
; Sequence 11, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/290,919
;       FILING DATE:  04-OCT-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9203821.5
;       FILING DATE:  22-FEB-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/GB93/00367
;       FILING DATE:  22-FEB-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KOKULIS, PAUL N.
;       REGISTRATION NUMBER:  16,773
;       REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 861-3000
;       TELEFAX:  (202) 822-0944
;       TELEX:  6714627 CUSH
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  106 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-290-919-11


  Query Match           12.5%;  Score 255.5;  DB 1;  Length 106;
  Best Local Similarity   47.1%;  Pred. No. 2.8e-12;
  Matches   48;  Conservative   19;  Mismatches   32;  Indels    3;  Gaps    3;

Qy    278 DMLNIS-QHQCV-KKQCPENSGCFRHLDEREECKCLLNYKQ-EGDKCEENPNPTCNENNG 334
          |:| :  :| ||  :  |:|:||||  :  || :||| ||: ||: | || ||||: |||
Db      5 DLLGVDPKHVCVDTRDIPKNAGCFRDDNGTEEWRCLLGYKKGEGNTCVENNNPTCDINNG 64

Qy    335 GCDADATCTEEDSGSSRKKITCECTKPDSYPLFDGIFCSSSN 376
          |||  |:|   :|  : ||| | | :|     ::|:|||||:
Db     65 GCDPTASCQNAESTENSKKIICTCKEPTPNAYYEGVFCSSSS 106


RESULT   7
US-08-257-073-5
; Sequence 5, Application US/08257073
; Patent No. 5766597
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  de Taisne, Charles
;    APPLICANT:  Tine, John A.
;    TITLE OF INVENTION:  MALARIA RECOMBINANT POXVIRUS VACCINE
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue, 25th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/257,073
;      FILING DATE:  09-JUN-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/075,783
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/852,305
;      FILING DATE:  18-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,183
;      FILING DATE:  20-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2570
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066 CURTMS
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  740 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-257-073-5


  Query Match           6.2%;  Score 127.5;  DB 1;  Length 740;
  Best Local Similarity   21.8%;  Pred. No. 0.1;
  Matches   95;  Conservative   58;  Mismatches  114;  Indels  169;  Gaps   22;

Qy     39 NIITFNLNL---NDILNSRLKKR---KYFLDVLESDLMQFKHISSNEYI------IEDSF 86
          |   : | :   ||:||  :|       |:||::| | |   : ||: |       |
Db    278 NATAYELGIVPVNDVLNVNMKNMITGNNFMDVVKSTLAQSGGLGSNDLINFLNQGKEIGE 337

Qy     87 KLLNSEQKNI---------------LLK----SYKYI----KESVENDIK------FAQE 117
           |||  : |:                :||    :|::|    | || | :|       :
Db    338 NLLNITKMNLGDKNNLESFPLDQLNMLKDNLINYEFILNNLKTSVLNKLKDLLLRLLYKA 397

Qy    118 GISY-----YEKVL-------AKYKDDLE-------------SIKKVIKEEKEK-FPSSP 151
           :||      || |         :| ::|:                 :| ::|: | | ||
Db    398 YVSYKKRKAQEKGLPEPTVTNEEYVEELKKGILDMGIKLLFSKVKSLLKKLKNKIFPKKK 457

Qy    152 PTTPPSPAKTDEQKKESKFLPFL--------TNIETLYNNLVNKIDDYLINLKAKINDCN 203
                |: |: |  |  | |         :||    ||:::|| :   |    |  |
Db    458 EDNQAVDTKSMEEPK-VKAQPALRGVEPTEDSNIMNSINNVMDEIDFFEKELIENNNTPN 516

Qy    204 V-----------------------------------EKDEAHVKITKLS-DLKAIDDK 225
          |                                   | |:  ||: |:   ||    |
Db    517 VVPPTQSKKKNKNETVSGMDENFDNHPENYFKEEYYYDENDDMEVKVKKIGVTLK----K 572

Qy    226 IDLFKNTNDFEAIK------------KLINDDTK------KDMLGKLL-----STGLVQI 262
           :  || |  | ||               : ||:| :      : :  :|:     :  : ||
Db    573 FEPLKNGNVSETIKLIHLGNKDKKHIEAINNDIQIIKQELQAIYNELMNYTNGNKNIQQI 632

Qy    263 FPNTIISKLIEGKFQDMLNISQHQCVKKQ---------------CPENSGCFRHLDEREE 307
          |   |:         |:||    : ::||                |:|       :| |:
Db    633 FQQNILE-------NDVLNQETEEEMEKQVEAITKQIEAEVDALAPKNK------EEEEK 679

Qy    308 CKCLLNYKQEGDKCEE 323
          |     |:| :| |:
Db    680 EKEKEKEKEEKEKEEK 695


RESULT   8
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2710 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-6


  Query Match             6.2%;  Score 127.5;  DB 1;  Length 2710;
  Best Local Similarity   23.4%;  Pred. No. 0.52;
  Matches   74;  Conservative   47;  Mismatches  102;  Indels   93;  Gaps   17;

Qy      36 IEKNIITFNLNLNDI-LNSRLKKRKYFL----------DVLESDLMQFKHISSNEYIIED 84
           : || ||   |  :: :||  : ||  |          : : |||      || |||  |
Db     732 VNKNSITIGANQYEVRINS--EGRKELLAHSGKWINKEEAIMSDL------SSKEYIFFD 783

Qy      85 SF-KLLNSEQKNI--LLKSYKYIK-----ESVENDIKFAQEGIS----------YYEK- 124
           |    | :: |||  |    : ||      ||  | ||    :            ||||
Db     784 SIDNKLKAKSKNIPGLASISEDIKTLLLDASVSPDTKFILNNLKLNIESSIGDYIYYEKL 843

Qy     125 ------------------VLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPA------ 159
                            :|    |:|  :||:   :::   |    :  :
Db     844 EPVKNIIHNSIDDLIDEFNLLENVSDELYELKKLNNLDEKYLISFEDISKNNSTYSVRFI 903

Qy     160 -KTD------EQKKE--SKFLPFLT-NIETLYNNLVNKIDDYLINLKAKINDCNVEKDEA 209
            |::      | :||  ||:   :|  | |: |:::  ::  |::        |:: |
Db     904 NKSNGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGNLLD--------NIQLDHT 955

Qy     210 HVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLIND---DTKKDMLGKLLSTGLVQIFPNT 266
             ::  |:    |  ||  |       | ::||     |  :  :| ||||  |: :
Db     956 S-QVNTLNAAFFIQSLIDYSSN-------KDVLNDLSTSVKVQLYAQLFSTGLNTIYDSI 1007

Qy     267 IISKLIEGKFQDMLNI 282
           :  ||     | :|:
Db    1008 QLVNLISNAVNDTINV 1023


RESULT   9
US-08-405-496A-6
; Sequence 6, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2710 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-6


  Query Match             6.2%;  Score 127.5;  DB 2;  Length 2710;
  Best Local Similarity   23.4%;  Pred. No. 0.52;
  Matches   74;  Conservative   47;  Mismatches  102;  Indels   93;  Gaps   17;

Qy      36 IEKNIITFNLNLNDI-LNSRLKKRKYFL----------DVLESDLMQFKHISSNEYIIED 84
           : || ||   |  :: :||  : ||  |          : : |||      || |||  |
Db     732 VNKNSITIGANQYEVRINS--EGRKELLAHSGKWINKEEAIMSDL------SSKEYIFFD 783

Qy      85 SF-KLLNSEQKNI--LLKSYKYIK-----ESVENDIKFAQEGIS----------YYEK- 124
           |    | :: |||  |    : ||      ||  | ||    :            ||||
Db     784 SIDNKLKAKSKNIPGLASISEDIKTLLLDASVSPDTKFILNNLKLNIESSIGDYIYYEKL 843

Qy     125 ------------------VLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPA------ 159
                            :|    |:|  :||:   :::   |    :  :
```

```
Db    844 EPVKNIIHNSIDDLIDEFNLLENVSDELYELKKLNNLDEKYLISFEDISKNNSTYSVRFI 903

Qy    160 -KTD------EQKKE--SKFLPFLT-NIETLYNNLVNKIDDYLINLKAKINDCNVEKDEA 209
Db    904 NKSNGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGNLLD--------NIQLDHT 955

Qy    210 HVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLIND---DTKKDMLGKLLSTGLVQIFPNT 266
Db    956 S-QVNTLNAAFFIQSLIDYSSN-------KDVLNDLSTSVKVQLYAQLFSTGLNTIYDSI 1007

Qy    267 IISKLIEGKFQDMLNI 282
Db   1008 QLVNLISNAVNDTINV 1023


RESULT 10
US-07-728-215-30
; Sequence 30, Application US/07728215
; Patent No. 5962643
;  GENERAL INFORMATION:
;    APPLICANT:  Sheppard, Dean
;    APPLICANT:  Quaranta, Vito
;    APPLICANT:  Pytela, Robert
;    TITLE OF INVENTION:  A No. 5962643el Integrin Beta Subunit and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/728,215
;      FILING DATE:  19910711
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P31 8717
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  798 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-728-215-30


  Query Match             6.2%;  Score 126.5;  DB 2;  Length 798;
  Best Local Similarity   21.0%;  Pred. No. 0.13;
  Matches  105;  Conservative   69;  Mismatches  137;  Indels  189;  Gaps   32;

Qy     48 NDILNSR----LKKRKYFLDVLESDLMQFK-----HISSNEYIIEDSFKLLNSEQKNILL 98
Db     78 DDIENPRGSKDIKKNKNVTNRSKGTAEKLKPEDIHQIQPQQLVL----RLRSGEPQTFTL 133

Qy     99 KSYKYIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKV---IKEEKEKFPSS----- 150
Db    134 KFKR--AEDYPIDLYYLMD-LSY------SMKDDLENVKSLGTDLMNEMRRITSDFRIGF 184

Qy    151 --------PPTTPPSPAK-----TDEQKKESKF----LPFLTNIETLYNNLVNK------ 187
Db    185 GSFVEKTVMPYISTTPAKLRNPCTSEQNCTTPFSYKNVLSLTNKGEVFNELVGKQRISGN 244

Qy    188 ID------DYLINL----------------------------KAKI------ND--CNVE 205
Db    245 LDSPEGGFDAIMQVAVCGSLIGWRNVTRLLVFSTDAGFHFAGDGKLGGIVLPNDGQCHLE 304

Qy    206 KDE------------AHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKKDMLGK 253
Db    305 NNMYTMSHYYDYPSIAHL-VQKLSE----NNIQTIFAVTEEFQPVYKELKNLIPKSAVGT 359

Qy    254 LL--STGLVQIF---PNTIISKLI--EGKFQDMLNISQHQCVK---------------- 289
Db    360 LSANSSNVIQLIIDAYNSLSSEVILENGKLSEGVTISYKSYCKNGVNGTGENGRKCSNIS 419

Qy    290 -------------KQCP-ENSGCFR-----HLDERE-------ECKCLLNYKQEGDKCEE 323
Db    420 IGDEVQFEISITSNKCPKKDSDSFKIRPLGFTEEVEVILQYICECECQSEGIPESPKCHE 479

Qy    324 NPNPT-------CNENNGG----C--------DADATCTEEDSG---SSRKKIT---CEC 358
Db    480 G-NGTFECGACRCNEGRVGRHCECSTDEVNSEDMDAYCRKENSSEICSNNGECVCGQCVC 538

Qy    359 TKPD-SYPLFDGIFCSSSNF 377
Db    539 RKRDNTNEIYSGKFCECDNF 558


RESULT 11
US-08-755-587-186
; Sequence 186, Application US/08755587
; Patent No. 6045997
;  GENERAL INFORMATION:
;    APPLICANT:  Futreal, Phillip A
;    APPLICANT:  Wooster, Richard F
;    APPLICANT:  Ashworth, Alan
;    APPLICANT:  Stratton, Michael R
;    TITLE OF INVENTION:  Materials and methods relating to the
;    TITLE OF INVENTION:  identification and sequencing of the BRCA2 cancer
;    TITLE OF INVENTION:  susceptibility gene and uses thereof.
;    NUMBER OF SEQUENCES:  222
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson
;      STREET:  310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
;      CITY:  Raleigh
;      STATE:  NC
;      COUNTRY:  USA
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/755,587
;      FILING DATE:  25-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9523959.6
;      FILING DATE:  23-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9525555.0
;      FILING DATE:  14-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9617961.9
;      FILING DATE:  28-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kenneth D Sibley
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5405-135
;  INFORMATION FOR SEQ ID NO:  186:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1494 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-755-587-186
```

```
  Query Match             5.9%;  Score 120.5;  DB 3;  Length 1494;
  Best Local Similarity   22.6%;  Pred. No. 0.81;
  Matches   94;  Conservative   68;  Mismatches  167;  Indels   87;  Gaps   21;

Qy       4 VTMDNILSGFEN-----EYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLN---DILNSR 54
Db     760 VTMSNKQQLTANTGQNIKDFDTFYL--------SFQTASRKNIRVSRESLNKARSLLNQK 811

Qy      55 LKKRKY--FLDVLESDLMQFKHIS----SNEYIIEDSFKLLNSEQK-----------NI 96
Db     812 WTEEELNNFSDSLNSELLPGIDIKKTDISNHEVIE------NTERKDKITKESLIGTENI 865

Qy      97 LLKSYKYIKESVENDIKFAQEG--ISYYE---KVLAKYKDDLESIKKVIKEEKEKFPSSP 151
Db     866 LL----ILQQRPESKIKKIKESAVLGFHTASGKKIEITKESLDKVKNLF-EEKEQDNSEI 920

Qy     152 PTTPPSPAKTDEQKKESK----------FLPFLTNIETLYNNL-VNKIDDYLINLKAKIN 200
Db     921 TNFSHRGAKMSKDREECKDGRELACGTTEITTTPEYEETHSSLEKKKVSNEIAALRPRLL 980

Qy     201 DCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKKDMLGKLLST-GL 259
Db     981 SDNLYKQTENLKISDHA-----SQKVDVHENTE---------KETAKKPTMYTNQSTYSA 1026

Qy     260 VQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGD 319
Db    1027 IENSPLT-FTQDTEEKF----SVSEASLFEAKKWLREGEWDDQSERINA-AKVNCKEYPD 1080

Qy     320 KCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDSYPLFDGIFCSSS 375
Db    1081 DYVENPSCGNSSNSAITENDKNHLSEKQGST---YLSNSTMSNSYSYHPG-FCHSS 1132
```

RESULT 12
US-09-213-053-2
; Sequence 2, Application US/09213053
; Patent No. 6159477
; GENERAL INFORMATION:
; APPLICANT: AUDONNET, Jean-Christophe
; APPLICANT: BAUDU, Philippe
; TITLE OF INVENTION: CANINE HERPESVIRUS BASED RECOMBINANT LIVE VACCINE, IN
; TITLE OF INVENTION: PARTICULAR AGAINST CANINE DISTEMPER, RABIES OR THE
; TITLE OF INVENTION: PARAINFLUENZA 2 VIRUS
; FILE REFERENCE: 454313-2200
; CURRENT APPLICATION NUMBER: US/09/213,053
; CURRENT FILING DATE: 1998-12-16
; EARLIER APPLICATION NUMBER: 9608242
; EARLIER FILING DATE: 1996-06-27
; EARLIER APPLICATION NUMBER: PCT/FR97/01115
; EARLIER FILING DATE: 1997-06-23
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1394
; TYPE: PRT
; ORGANISM: Canine herpesvirus
US-09-213-053-2

```
  Query Match             5.8%;  Score 118;  DB 4;  Length 1394;
  Best Local Similarity   19.2%;  Pred. No. 1.1;
  Matches  102;  Conservative   65;  Mismatches  153;  Indels  212;  Gaps   25;

Qy      18 DVIYLKPLAGVYRSLK---------KQIEKNIITFNLNLNDILNSRLKKRKYF--LDV-- 64
Db     866 DTTITKPISTEANNLKSVSTSIKPPKNLKKKLLKSKCRDNVIY------RPYFSQLEINC 919

Qy      65 -------LESDLMQ--FKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVEND---- 111
Db     920 TITKKQNLSNPLIELWFKELSTYNKTNENVESLKTDISKNILLFSTKNNSDNFYNDFLLG 979

Qy     112 -----IKFAQEGISYYE--KVLAK----------------YKDDLESIKKVIKEEKEKFP 148
Db     980 IQNQPVNYKLYGSQFYDNGNILLNIKSVDFKTSGIYTWKLYNSNNESIFETFKIQVYAYH 1039

Qy     149 SSPPTTPPSPAKTDEQKKE----SKFLPFLTNIETLYNNLVNKIDDYLINLKAKINDCNV 204
Db    1040 SPNVNLKSNPSLYNENYSAICTIANYFP-LESTEIFWFNDGQPIDKKYI----------- 1087

Qy     205 EKDEAH----------VKITKLSDLKAIDD------KIDLFKNTNDFEAIKKLINDDTKK 248
Db    1088 --DETYSVWIDGLITRTSILSLPFSEAMESPPNLRCNVEWYKNS---KASKK-------- 1134

Qy     249 DMLGKLLSTGLVQIFPNTIISKL-----IEGKFQDMLNISQHQCVK----------KQCP 293
Db    1135 -------------FSNTVIPKVYYKPFISIKFDNGLAICDAKCVSRENNKLQWLVKDIP 1180

Qy     294 EN-----SG-CFRH-------------LDEREECKCLL--------NYKQEG--DKCEE 323
Db    1181 INGDDIISGPCLNHPGLVNIQNKIDISDYDEPVTYKCSIIGYPIIFPNFYDEKVFDASDE 1240

Qy     324 NPNPT---------------------------CNENNGGC-------------DADATC 342
Db    1241 NVSKSMLISITTIIGGAIFVIVLIFITALCFYCSKNNKIMAQLVLTDIPLEDVENKNTSS 1300

Qy     343 TEEDSGSSRKKITCECTKPDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
Db    1301 DEETTNLNQKKSTCQC-----------LCVTLGFFAAGILLTIAAIIFTFI 1340
```

RESULT 13
US-08-286-325A-2
; Sequence 2, Application US/08286325A
; Patent No. 5658770
; GENERAL INFORMATION:
; APPLICANT: PREVOTS, Fabien
; APPLICANT: REMY, Elisabeth
; APPLICANT: RITZENTHALER, Paul
; TITLE OF INVENTION: Nucleic acid sequence and plasmids
; TITLE OF INVENTION: comprising at least one phage resistance mechanism,
; TITLE OF INVENTION: bacteria in which they are present, and their use
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BACON & THOMAS
; STREET: 625 Slaters Lane - Fourth Floor
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,325A
; FILING DATE: 04-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93 09 777
; FILING DATE: 09-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FICHTER, Richard E
; REGISTRATION NUMBER: 26,382
; REFERENCE/DOCKET NUMBER: REF/BEDL/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 683-0500
; TELEFAX: (703) 683-1080
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 540 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-286-325A-2

```
  Query Match             5.7%;  Score 116;  DB 1;  Length 540;
  Best Local Similarity   25.0%;  Pred. No. 0.48;
  Matches   70;  Conservative   45;  Mismatches   89;  Indels   76;  Gaps   18;

Qy      28 VYRSLKKQIEKN---IITFNLNLNDILNSRLKKRKYFLDVLESDLMQFKHISSNEYIIED 84
           ||   || :|||    |    |||:    ||  | : :  ::| :    :    :||: |
Db     250 VYHICKKLMEKNKRVAIVHGANLNN-GQQRLALRGFTIFPVKSIIEVLDNADKYDYIVVD 308

Qy      85 SFKLLNSEQKNILLKSYKYIKESVEN-DIKF--AQEGISYYEKVLAKYKDDLESIK--KV 139
             : |  :    | : |  : :::||   ||  : :|     : | ||: :  |||  |
Db     309 EAQRLRQD----LGEQYTKLVDTIENSQTKFIISLDG----RQTLNKYEIEENSIKLFKY 360

Qy     140 IKEE------KEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYN--NLVNKIDDY 191
           || :      |:|| ::|              : |||:  |  |  :|   :|:: ||
Db     361 IKNKGVTFSLKDKFRTNP-------------EMSKFIQLLFKI-PMYKKIDLISNIDH- 404

Qy     192 LINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKKDML 251
             |:  |  | | |    ::     ||:          : :|:| :     |  :|  :
Db     405 --NIIIKYFD-NRESGNEYI-----SDM----------DSNSDWEVL-NYTKDRFRKTGI 445

Qy     252 GKLLSTGLV------QIFPNTII----------SKLIEGK 275
           ||:   ||       | |   ||           |:|: |
Db     446 GKMCGNGLTSHSIIGQEFDKVIIPLDSNFFYKEQKIIDSK 485


RESULT  14
US-08-480-640A-115
; Sequence 115, Application US/08480640A
; Patent No. 6033904
;  GENERAL INFORMATION:
;    APPLICANT:  Cochran, Mark D.
;    APPLICANT:  Junker, David E.
;    TITLE OF INVENTION:  Recombinant Swinepox Virus
;    NUMBER OF SEQUENCES:  225
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,640A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  115:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  677 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Swinepox virus
;      STRAIN:  Kasza
;      INDIVIDUAL ISOLATE:  S-SPV-001
;    IMMEDIATE SOURCE:
;      CLONE:  515-85.1
;    POSITION IN GENOME:
;      MAP POSITION:  ~23.2
;      UNITS:  %G
US-08-480-640A-115


  Query Match             5.6%;  Score 115.5;  DB 3;  Length 677;
  Best Local Similarity   21.6%;  Pred. No. 0.69;
  Matches   71;  Conservative   59;  Mismatches  123;  Indels   75;  Gaps   16;

Qy      22 LKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLE---SDLMQFKHISSN 78
           |: :  :   |||    |||        | ::  :|    |: :|    |::       | |
Db     199 LRDINRIIELLKKYPNNNII-------DYISDSIKSNSSFIHILHMIISNMFPAIIPSVN 251

Qy      79 EYI--IEDSFKLLNSEQ-KNILLKSYKY----IKESVENDIKFAQEGISYYEKVLAKYKD 131
           ::|  : |  :|:|    | : : ||      ::   ::|  | :: || |:    | :|
Db     252 DFISTVVDKDRLINMYGIKCVAMFSYDINMIDLESLDDSDYIFIEKNISIYD---VKCRD 308

Qy     132 DLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKIDDY 191
             |:  :| || :  ::         | ::  :  |        :   |:  ||:::
Db     309 FANMIRDKVKREKNRILTT---------KCEDIIRYIKLFS-----KNRINDENNKVEEV 354

Qy     192 LINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKKDML 251
           ||::       || |:        |||: :: |:  |    |       |::  | |
Db     355 LIHID------NVSKNNK----LSLSDISSLMDQFRL----NPCTIRNILLSSATIK--- 397

Qy     252 GKLLSTGLVQ-----IFPNTIISKLIEGKF-QDMLNISQHQCVK--KQCPENSGCFRHLD 303
           |||:   |:        |  : | |:|    ||::      :|  ||  :    | :
Db     398 SKLLALRAVKNWKCYSLTNVSMYKKIKGVIVMDMVDYISTNILKYHKQLYDKMSTFEYKR 457

Qy     304 EREECKC---------------LLNYK 315
           : : |||                 :|||
Db     458 DIKSCKCSICSDSITHHIYETTSCINYK 485


RESULT  15
US-08-480-640A-193
; Sequence 193, Application US/08480640A
; Patent No. 6033904
;  GENERAL INFORMATION:
;    APPLICANT:  Cochran, Mark D.
;    APPLICANT:  Junker, David E.
;    TITLE OF INVENTION:  Recombinant Swinepox Virus
;    NUMBER OF SEQUENCES:  225
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,640A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  193:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  677 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-640A-193
```

Query Match 5.6%; Score 115.5; DB 3; Length 677;
Best Local Similarity 21.6%; Pred. No. 0.69;
Matches 71; Conservative 59; Mismatches 123; Indels 75; Gaps 16;

```
Qy      22 LKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLE---SDLMQFKHISSN 78
           |: :  :   |||    |||       | ::  :|    |: :|    |::      | |
Db     199 LRDINRIIELLKKYPNNNII-------DYISDSIKSNSSFIHILHMIISNMFPAIIPSVN 251

Qy      79 EYI--IEDSFKLLNSEQ-KNILLKSYKY----IKESVENDIKFAQEGISYYEKVLAKYKD 131
           ::|  : |  :|:|    | : : ||      ::   ::|  | :: || |:    | :|
Db     252 DFISTVVDKDRLINMYGIKCVAMFSYDINMIDLESLDDSDYIFIEKNISIYD---VKCRD 308

Qy     132 DLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKIDDY 191
               |:  :| || :  ::         | ::  :  |        :   |:  ||:::
Db     309 FANMIRDKVKREKNRILTT---------KCEDIIRYIKLFS-----KNRINDENNKVEEV 354

Qy     192 LINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKKDML 251
           ||::       || |:        |||: :: |:  |    |       |::  | |
Db     355 LIHID------NVSKNNK----LSLSDISSLMDQFRL----NPCTIRNILLSSATIK--- 397

Qy     252 GKLLSTGLVQ-----IFPNTIISKLIEGKF-QDMLNISQHQCVK--KQCPENSGCFRHLD 303
            |||:   |:        |  : | |:|    ||::      :|  ||  :    | :
Db     398 SKLLALRAVKNWKCYSLTNVSMYKKIKGVIVMDMVDYISTNILKYHKQLYDKMSTFEYKR 457

Qy     304 EREECKC----------------LLNYK 315
           : : |||                :|||
Db     458 DIKSCKCSICSDSITHHIYETTSCINYK 485
```

Search completed: August 8, 2001, 12:29:58
Job time: 46 sec

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:31:53 ; Search time 60.73 Seconds
(without alignments)
393.312 Million cell updates/sec

Title: US-09-500-376-3
Perfect score: 2052
Sequence: 1 AISVTMDNILSGFENEYDVI.....SNFLGISFLLILMLILYSFI 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1248.5	60.8	1639	19	AAW54145	P. falciparum synt
2	1156.5	56.4	355	20	AAY09372	Merozoite surface
3	1156.5	56.4	355	20	AAY05832	Merozoite surface
4	1156.5	56.4	361	20	AAY09373	Merozoite surface
5	1156.5	56.4	361	20	AAY05833	Merozoite surface
6	1149.5	56.0	376	20	AAY09374	Modified merozoite
7	1149.5	56.0	376	20	AAY05834	Modified merozoite
8	1091	53.2	1654	6	AAP50777	Sequence of the P1
9	641.5	31.3	262	22	AAB37610	Merozoite surface
10	615	30.0	116	18	AAW36103	PfMSP1(p19)A prote
11	615	30.0	116	18	AAW22592	PfMSP1(p19)A prote
12	529	25.8	108	22	AAB37609	Merozoite surface
13	528	25.7	96	22	AAB37608	Merozoite surface
14	521	25.4	127	18	AAW22593	PfMSP1(p19)S prote
15	521	25.4	127	18	AAW36102	PfMSP1(p19)S prote
16	304	14.8	54	14	AAR41357	MSP1EGF2B EGF2-lik
17	288	14.0	54	14	AAR41356	MSP1EGF2A EGF2-lik
18	276	13.5	49	14	AAR41355	MSP1EGF1B EGF1-lik
19	273	13.3	49	14	AAR41354	MSP1EGF1A EGF1-lik
20	271	13.2	93	22	AAB37611	Merozoite surface
21	255.5	12.5	106	14	AAR41358	P. yoelii combined
22	187.5	9.1	350	21	AAY70278	Recombinant vaccin
23	166	8.1	980	21	AAB18294	Plasmodium falcipa
24	145.5	7.1	1712	21	AAB18205	Plasmodium falcipa
25	139	6.8	1979	21	AAB18171	Plasmodium falcipa
26	134	6.5	2573	21	AAB18234	Plasmodium falcipa
27	134	6.5	4134	20	AAY31946	Plasmodium falcipa
28	132.5	6.5	1188	21	AAB18183	Plasmodium falcipa
29	132.5	6.5	1346	21	AAB18236	Plasmodium falcipa
30	132	6.4	1166	21	AAB18268	Plasmodium falcipa
31	129.5	6.3	1817	21	AAB18255	Plasmodium falcipa
32	128.5	6.3	646	21	AAB18188	Plasmodium falcipa
33	128.5	6.3	2539	21	AAB18198	Plasmodium falcipa
34	127.5	6.2	740	13	AAR27530	Plasmodium falcipa
35	127.5	6.2	740	16	AAR68838	Plasmodium falcipa
36	127.5	6.2	2710	17	AAR95016	C. difficile toxin
37	127.5	6.2	2710	19	AAW68387	Clostridium diffic
38	127	6.2	26	22	AAB71974	Plasmodium falcipa
39	127	6.2	2500	21	AAB18272	Plasmodium falcipa
40	126.5	6.2	798	22	AAB36937	Human integrin bet
41	126.5	6.2	963	19	AAW70540	Integrin beta-1 ch
42	126.5	6.2	1743	19	AAW98879	H. pylori GHPO 175
43	126	6.1	2285	20	AAW98149	Bacillus subtilis
44	124.5	6.1	389	19	AAW98813	H. pylori GHPO 137
45	123.5	6.0	717	21	AAB21231	Tomato LeMPF1. Ly

ALIGNMENTS

RESULT 1
AAW54145
ID AAW54145 standard; Protein; 1639 AA.
XX
AC AAW54145;
XX
DT 23-SEP-1998 (first entry)
XX
DE P. falciparum synthetic gp190 protein.
XX
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW monoclonal antibody; passive immunisation; parasite.
XX
OS Plasmodium falciparum.
OS Synthetic.
XX
PN WO9814583-A2.
XX
PD 09-APR-1998.
XX
PF 02-OCT-1997; 97WO-EP05441.
XX
PR 02-OCT-1996; 96DE-4040817.
XX
PA (BUJA/) BUJARD H.
XX
PI Bujard H, Pan W, Tolle R;
XX
DR WPI; 1998-240088/21.
DR N-PSDB; AAV21451, AAV35363.
XX
PT Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PT reducing their AT content

```
XX
PS   Example 1; Fig 3c; 48pp; German.
XX
CC   This sequence represents a modified Plasmodium falciparum gp190/MSP-1
CC   (merozoite surface) protein. The gene encoding this protein has been
CC   stabilised by reducing the AT content of the nucleotide sequence. Such a
CC   protein is useful in vaccines against malaria or for producing monoclonal
CC   antibodies (for passive immunisation). The complete gp190 protein can now
CC   be produced outside the parasite and has, at least over extended regions,
CC   the native pattern of folding. Larger amounts of the protein can be
CC   produced recombinantly than would be possible using the parasites as
CC   source.
XX
SQ   Sequence   1639 AA;


  Query Match             60.8%;  Score 1248.5;  DB 19;  Length 1639;
  Best Local Similarity   61.0%;  Pred. No. 3e-79;
  Matches  239;  Conservative   54;  Mismatches   74;  Indels   25;  Gaps    3;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db    1270 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 1329

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db    1330 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 1389

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db    1390 sekyksdldsikkyi--------------------ndkqgenekylpflnnietlyktvn 1429

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
           :||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db    1430 dkidlfvihleakvlnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 1489

Qy     246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHL 302
             ::| | ||||:|  | |   |::| |::|  | ||||||||||||||:||||||||
Db    1490 nhnnlltkflstgmv--fenlaktvlsnlldgnlqgmlnisqhqcvkkqcpqnsgcfrhl 1547

Qy     303 DEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPD 362
           |||||||||||||||||| |||||||||||||||| |||||||: ||||||||||
Db    1548 dereeckcllnykqegdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpd 1607

Qy     363 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||| ||||||||||||
Db    1608 syplfdgifcsssnflgiffllilmlilysfi 1639


RESULT   2
AAY09372
ID   AAY09372 standard; Protein; 355 AA.
XX
AC   AAY09372;
XX
DT   31-AUG-1999  (first entry)
XX
DE   Merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal.
XX
OS   Plasmodium falciparum.
XX
PN   WO9920774-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22226.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-288313/24.
DR   P-PSDB; AAX56008.
XX
PT   Modified malarial  protein for use in anti-malarial vaccines
XX
PS   Example; Fig 1; 35pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The nucleic acid sequence encoding MSP-1-42
CC   has been modified (see AAX56008) compared to the native sequence (see
CC   AAX56009) such that 306 nucleotide positions have been replaced to
CC   lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
CC   instability motifs while maintaining the same protein amino acid
CC   sequence.  These alterations allow MSP-1-42 to be expressed in
CC   mammalian cell culture and in transgenic mice.  Native MSP-1-12
CC   is known to be difficult to express in cell culture systems,
CC   mammalian cell culture systems or in transgenic animals.  The
CC   invention allows expression of MSP-1 protein in the milk of
CC   transgenic animals, and also provides a DNA vaccine comprising a
CC   vector containing the altered MSP-1-42 sequence.
XX
SQ   Sequence   355 AA;


  Query Match             56.4%;  Score 1156.5;  DB 20;  Length 355;
  Best Local Similarity   59.0%;  Pred. No. 1.2e-73;
  Matches  219;  Conservative   54;  Mismatches   73;  Indels   25;  Gaps    3;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db       7 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 66

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db      67 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 126

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db     127 sekyksdldsikkyi--------------------ndkqgenekylpflnnietlyktvn 166

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
           :||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db     167 dkidlfvihleakvlnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 226

Qy     246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHL 302
             ::| | ||||:|  | |   |::| |::|  | ||||||||||||||:||||||||
Db     227 nhnnlltkflstgmv--fenlaktvlsnlldgnlqgmlnisqhqcvkkqcpqnsgcfrhl 284

Qy     303 DEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPD 362
           |||||||||||||||||| |||||||||||||||| |||||||: ||||||||||
Db     285 dereeckcllnykqegdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpd 344

Qy     363 SYPLFDGIFCS 373
           |||||||||||
Db     345 syplfdgifcs 355


RESULT   3
AAY05832
ID   AAY05832 standard; Protein; 355 AA.
XX
AC   AAY05832;
XX
DT   02-AUG-1999  (first entry)
XX
DE   Merozoite surface protein MSP-1-42.
```

XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX
OS Plasmodium falciparum.
XX
PN WO9920766-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22225.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI; 1999-302742/25.
DR N-PSDB; AAX25586.
XX
PT New modified recombinant nucleic acid sequences useful for producing
PT malarial DNA vaccine
XX
PS Disclosure; Fig 1; 43pp; English.
XX
CC The present sequence represents a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
CC has been modified (see AAX25586) compared to the native sequence (see
CC AAX25587) such that 306 nucleotide positions have been replaced to
CC lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
CC instability motifs while maintaining the same protein amino acid
CC sequence. These alterations allow MSP-1-42 to be expressed in
CC mammalian cell culture and in transgenic mice. The invention
CC provides modified recombinant nucleic acid sequences and methods for
CC increasing the mRNA levels and protein expression of proteins that
CC are difficult to express in cell culture systems, mammalian cell
CC culture systems or in transgenic animals. The preferred difficult
CC protein candidates for expression are those derived from lower
CC organisms such as parasites, bacteria and viruses that have DNA
CC coding sequences of high AT content or which have mRNA instability
CC motifs or rare codons relative to the recombinant expression system
CC to be used. The invention allows expression of MSP-1 protein in
CC the milk of transgenic animals, and also provides a DNA vaccine
CC comprising a vector containing the altered MSP-1-42 sequence.
XX
SQ Sequence 355 AA;

Query Match 56.4%; Score 1156.5; DB 20; Length 355;
Best Local Similarity 59.0%; Pred. No. 1.2e-73;
Matches 219; Conservative 54; Mismatches 73; Indels 25; Gaps 3;

```
Qy      6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
          :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db      7 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 66

Qy     66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
          ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db     67 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 126

Qy    126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
            ||| ||:|||| |              |:| :  |:|||| ||||||  :
Db    127 sekyksdldsikkyi----------ndkqgenekylpflnnietlyktvn 166

Qy    186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
          :||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db    167 dkidlfvihleakvlnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 226
```

Qy 246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHL 302
::| | ||||:| | | |::| |::| | ||||||||||||||||:||||||||
Db 227 nhnnlltkflstgmv--fenlaktvlsnlldgnlqgmlnisqhqcvkkqcpqnsgcfrhl 284

Qy 303 DEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSSRKKITCECTKPD 362
|||||||||||||||||| |||||||||||||||| ||||||||: ||||||||||
Db 285 dereeckcllnykqegdkcvenpnptcnenngqcdadakcteedsgsngkkitcectkpd 344

Qy 363 SYPLFDGIFCS 373
|||||||||||
Db 345 syplfdgifcs 355

RESULT 4
AAY09373
ID AAY09373 standard; Protein; 361 AA.
XX
AC AAY09373;
XX
DT 31-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX
OS Plasmodium falciparum.
XX
PN WO9920774-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22226.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI; 1999-288313/24.
DR P-PSDB; AAX56009.
XX
PT Modified malarial protein for use in anti-malarial vaccines
XX
PS Example; Fig 2; 35pp; English.
XX
CC This present sequence comprises a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The C-terminal end of the sequence is
CC modified to include a 6xHis tag. A nucleic acid (see AAX56008)
CC encoding MSP-1-42 has been modified according to a method
CC of the invention in order to improve expression in mammalian cells
CC and in transgenic animals by reducing the AT content and removing
CC mRNA instability motifs. The invention allows expression of
CC MSP-1-42 in the milk of transgenic animals, and also provides a DNA
CC vaccine comprising a vector containing the altered MSP-1-42 nucleic
CC acid.
XX
SQ Sequence 361 AA;

Query Match 56.4%; Score 1156.5; DB 20; Length 361;
Best Local Similarity 59.0%; Pred. No. 1.2e-73;
Matches 219; Conservative 54; Mismatches 73; Indels 25; Gaps 3;

```
Qy      6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
          :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db      7 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 66
```

```
Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db      67 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 126

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db     127 sekyksdldsikkyi-------------------ndkqgenekylpflnnietlyktvn 166

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
           :||| ::|:|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db     167 dkidlfvihleakvlnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 226

Qy     246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHL 302
             ::| | ||||:|  | |   |::| |::|  | ||||||||||||||:||||||||
Db     227 nhnnlltkflstgmv--fenlaktvlsnlldgnlqgmlnisqhqcvkkqcpqnsgcfrhl 284

Qy     303 DEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPD 362
           |||||||||||||||||| |||||||||||||||| |||||||: ||||||||||
Db     285 dereeckcllnykqegdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpd 344

Qy     363 SYPLFDGIFCS 373
           ||||||||||
Db     345 syplfdgifcs 355
```

RESULT 5
AAY05833
ID AAY05833 standard; Protein; 361 AA.
XX
AC AAY05833;
XX
DT 02-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX
OS Plasmodium falciparum.
XX
PN WO9920766-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22225.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI; 1999-302742/25.
DR N-PSDB; AAX25587.
XX
PT New modified recombinant nucleic acid sequences useful for producing
PT malarial DNA vaccine
XX
PS Disclosure; Fig 2; 43pp; English.
XX
CC This present sequence comprises a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The C-terminal end of the sequence is
CC modified to include a 6xHis tag. Nucleic acids (see AAX25586 and
CC AAX25593) encoding MSP-1-42 have been modified according to a method
CC of the invention in order to improve expression in mammalian cells
CC and in transgenic animals. The invention provides modified
CC recombinant nucleic acid sequences and methods for increasing the
CC mRNA levels and protein expression of proteins that are difficult
CC to express in cell culture systems, especially mammalian cell
CC culture systems or in transgenic animals. The preferred difficult
CC protein candidates for expression are those derived from lower
CC organisms such as parasites, bacteria and viruses that have DNA
CC coding sequences of high AT content or which have mRNA instability
CC motifs or rare codons relative to the recombinant expression system
CC to be used. The invention allows expression of MSP-1 in the milk
CC of transgenic animals, and also provides a DNA vaccine comprising a
CC vector containing the altered MSP-1-42 nucleic acid.
XX
SQ Sequence 361 AA;

Query Match 56.4%; Score 1156.5; DB 20; Length 361;
Best Local Similarity 59.0%; Pred. No. 1.2e-73;
Matches 219; Conservative 54; Mismatches 73; Indels 25; Gaps 3;

```
Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
             :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db       7 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 66

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db      67 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 126

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db     127 sekyksdldsikkyi-------------------ndkqgenekylpflnnietlyktvn 166

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
           :||| ::|:|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db     167 dkidlfvihleakvlnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 226

Qy     246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHL 302
             ::| | ||||:|  | |   |::| |::|  | ||||||||||||||:||||||||
Db     227 nhnnlltkflstgmv--fenlaktvlsnlldgnlqgmlnisqhqcvkkqcpqnsgcfrhl 284

Qy     303 DEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPD 362
           |||||||||||||||||| |||||||||||||||| |||||||: ||||||||||
Db     285 dereeckcllnykqegdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpd 344

Qy     363 SYPLFDGIFCS 373
           ||||||||||
Db     345 syplfdgifcs 355
```

RESULT 6
AAY09374
ID AAY09374 standard; Protein; 376 AA.
XX
AC AAY09374;
XX
DT 31-AUG-1999 (first entry)
XX
DE Modified merozoite surface protein MSP-1-42.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal; mutant.
XX
OS Plasmodium falciparum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..15
FT /note= "beta-casein signal peptide"
FT Peptide 371..376
FT /note= "6xHis tag"
FT Misc-difference 197
FT /note= "Asn in native MSP-1-42 (N181Q mutation)"
FT Misc-difference 278
FT /note= "Asn in native MSP-1-42 (N262Q mutation)"
XX
PN WO9920774-A2.

```
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22226.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-288313/24.
DR   P-PSDB; AAX56008.
XX
PT   Modified malarial  protein for use in anti-malarial vaccines
XX
PS   Example; Fig 11; 35pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The sequence has been modified to include
CC   an N-terminal beta-casein signal peptide and an N-terminal 6xHis
CC   tag.  In addition, N181Q and N262Q mutations have been introduced to
CC   eliminate 2 N-glycosylation sites.  These modifications allow the
CC   MSP-1-42 protein to be expressed in the milk of transgenic mice.
XX
SQ   Sequence   376 AA;


  Query Match             56.0%;  Score 1149.5;  DB 20;  Length 376;
  Best Local Similarity   58.8%;  Pred. No. 4e-73;
  Matches  218;  Conservative   53;  Mismatches   75;  Indels   25;  Gaps    3;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db      22 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 81

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |::  | || |||:|:: || || : : ||: :
Db      82 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 141

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db     142 sekyksdldsikkyi-------------------ndkqgenekylpflnnietlyktvn 181

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
           :||| ::|:|:||:     ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db     182 dkidlfvihleakvlqytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 241

Qy     246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHL 302
             ::| | ||||:|  | |   |::| |::|  | || ||||||||||||:||||||||
Db     242 nhnnlltkflstgmv--fenlaktvlsnlldgnlqgmlqisqhqcvkkqcpqnsgcfrhl 299

Qy     303 DEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPD 362
           |||||||||||||||||| |||||||||||||||| ||||||||: ||||||||||
Db     300 dereeckcllnykqegdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpd 359

Qy     363 SYPLFDGIFCS 373
           ||||||||||
Db     360 syplfdgifcs 370


RESULT   7
AAY05834
ID   AAY05834 standard; Protein; 376 AA.
XX
AC   AAY05834;
XX
DT   02-AUG-1999  (first entry)
XX
DE   Modified merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal; mutant.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..15
FT                   /note= "beta-casein signal peptide"
FT   Peptide         371..376
FT                   /note= "6xHis tag"
FT   Misc-difference 197
FT                   /note= "Asn in native MSP-1-42 (N181Q mutation)"
FT   Misc-difference 278
FT                   /note= "Asn in native MSP-1-42 (N262Q mutation)"
XX
PN   WO9920766-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22225.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-302742/25.
DR   N-PSDB; AAX25593.
XX
PT   New modified recombinant nucleic acid sequences useful for producing
PT   malarial DNA vaccine
XX
PS   Example; Fig 11; 43pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The sequence has been modified to include
CC   an N-terminal beta-casein signal peptide and an N-terminal 6xHis
CC   tag.  In addition, N181Q and N262Q mutations have been introduced to
CC   eliminate 2 N-glycosylation sites.  These modifications allow the
CC   MSP-1-42 protein to be expressed in the milk of transgenic mice.
CC   The invention generally relates to modified recombinant nucleic
CC   acid sequences and methods for increasing the mRNA levels and
CC   protein expression of proteins that are difficult to express in
CC   cell culture systems, mammalian cell culture systems or in
CC   transgenic animals.  The preferred difficult protein candidates for
CC   expression are those derived from lower organisms such as parasites,
CC   bacteria and viruses that have DNA coding sequences of high AT
CC   content or which have mRNA instability motifs or rare codons
CC   relative to the recombinant expression system to be used.
XX
SQ   Sequence   376 AA;


  Query Match             56.0%;  Score 1149.5;  DB 20;  Length 376;
  Best Local Similarity   58.8%;  Pred. No. 4e-73;
  Matches  218;  Conservative   53;  Mismatches   75;  Indels   25;  Gaps    3;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db      22 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 81

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |::  | || |||:|:: || || : : ||: :
Db      82 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 141

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
```

```
          ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db    142 sekyksdldsikkyi-------------------ndkqgenekylpflnnietlyktvn 181

Qy    186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
          :||| ::|:|:||:     ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db    182 dkidlfvihleakvlqytyeksnvevkikelnylktiqdkladfknnnfvgiadlstdy 241

Qy    246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHL 302
            ::| | ||||:|  | |   |::| |::|  | || ||||||||||||:||||||||
Db    242 nhnnlltkflstgmv--fenlaktvlsnlldgnlqgmlqisqhqcvkkqcpqnsgcfrhl 299

Qy    303 DEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPD 362
          |||||||||||||||||| |||||||||||||||| ||||||||: ||||||||||
Db    300 dereeckcllnykqegdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpd 359

Qy    363 SYPLFDGIFCS 373
          ||||||||||
Db    360 syplfdgifcs 370


RESULT   8
AAP50777
ID   AAP50777 standard; Protein; 1654 AA.
XX
AC   AAP50777;
XX
DT   30-SEP-1991  (first entry)
XX
DE   Sequence of the P195 protein of Plasmodium falciparum.
XX
KW   Malaria vaccine; epitope; antigen; immunogen.
XX
OS   Plasmodium falciparum.
XX
PN   EP154454-A.
XX
PD   11-SEP-1985.
XX
PF   21-FEB-1985;   85EP-0301173.
XX
PR   26-SEP-1984;   84GB-0024340.
PR   22-FEB-1984;   84GB-0004692.
PR   21-FEB-1985;   85GB-0004429.
XX
PA   (WELL ) WELLCOME FOUNDATION LTD.
XX
PI   Holder A,  Sandhu J,  Odink K,  Lockyer M,  Riveros-Moreno V;
XX
DR   WPI; 1985-224845/37.
DR   N-PSDB; AAN50530.
XX
PT   Cloned DNA sequence encoding plasmodium falciparum protein -
PT   useful for expressing the protein for use in vaccines against
PT   malaria
XX
PS   Claim 6; Fig 1; 51pp; English.
XX
CC   The sequence encoding the P195  protein of Plasmodium falciparum
CC   (AAN50530) and a peptide comprising at least one of its epitopes
CC   (see AAP50777) are claimed. Also claimed is a vaccine for inducing
CC   immunity to malaria comprising the novel peptide or P195 or a
CC   peptide comprising at least one epitope when derived from the new
CC   DNA sequence, together with a carrier.
XX
SQ   Sequence   1654 AA;


  Query Match           53.2%;  Score 1091;  DB 6;  Length 1654;
  Best Local Similarity   55.2%;  Pred. No. 3.6e-68;
  Matches  217;  Conservative   60;  Mismatches   90;  Indels   26;  Gaps    4;

Qy      6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
          :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
```

```
Db   1284 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 1343

Qy     66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
          ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db   1344 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 1403

Qy    126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
            ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db   1404 sekyksdldsikkyi-------------------ndkqgenekylpflnnietlyktvn 1443

Qy    186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
          :||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db   1444 dkidlfvihleakvlnytyeksnvevkikelnylktiqdkladfknnnfvgiadlstdy 1503

Qy    246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRH 301
            ::| | ||||:|  | |   :::| |:: |  : : : :     |    ::|||||
Db   1504 nhnnlltkflstgmv--fenllksvlsnlldwklaryvkhfttpmrkktmiqqssgcfrh 1561

Qy    302 LDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKP 361
          |||||||||||||||| || || |||||||||||||| ||||||||: ||||||||||
Db   1562 ldereeckcllnykqegskcvensnptcnennggcdadakcteedsgsngkkitcectkp 1621

Qy    362 DSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          | |||  ||||||||||||||||||||||||||
Db   1622 dcyplsmvifcsssnflgisfllilmlilysfi 1654


RESULT   9
AAB37610
ID   AAB37610 standard; Protein; 262 AA.
XX
AC   AAB37610;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Merozoite surface protein-133.
XX
KW   Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS   Plasmodium falciparum.
XX
PN   WO200063245-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-GB01558.
XX
PR   20-APR-1999;   99GB-0009072.
PR   13-MAY-1999;   99US-0311817.
PR   25-MAY-1999;   99CA-2271451.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR   WPI; 2001-015762/02.
DR   N-PSDB; AAC68978.
XX
PT   Novel variants of the C-terminal fragment of Plasmodium merozoite
PT   surface protein-1, useful as vaccines for treating or preventing
PT   malaria -
XX
PS   Example 5; Fig 15; 126pp; English.
XX
CC   The present invention relates to non-natural variants of a C-terminal
CC   fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC   non-natural variants have reduced affinity for at least 1 antibody
CC   capable of blocking a second antibody that inhibits the proteolytic
CC   cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC   one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC   MSP-1_4_2, compared to natural MSP-1_1_9. The non-natural variants of the
CC   present invention are useful for immunising a mammal against malaria, and
CC   can be used to treat malaria. The present sequence is MSP-133 protein.
```

```
XX
SQ   Sequence   262 AA;


  Query Match             31.3%;  Score 641.5;  DB 22;  Length 262;
  Best Local Similarity   47.1%;  Pred. No. 1.3e-37;
  Matches  131;  Conservative   51;  Mismatches   71;  Indels   25;  Gaps    3;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db       7 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 66

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db      67 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 126

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db     127 sekyksdldsikkyi--------------------ndkqgenekylpflnnietlyktvn 166

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
           :||| ::|:|:||:     ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db     167 dkidlfvihleakvlqytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 226

Qy     246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKFQDML 280
             ::| | ||||:|  | |   |::| |::|  | ||
Db     227 nhnnlltkflstgmv--fenlaktvlsnlldgnlqgml 262


RESULT  10
AAW36103
ID   AAW36103 standard; Protein; 116 AA.
XX
AC   AAW36103;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)A protein sequence.
XX
KW   Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          1..95
FT                   /note= "amino acids derived from P. falciparum MSP1 p19
FT                        fragment"
FT   Region          96..116
FT                   /note= "glycosylphosphatidylinositol anchoring sequence"
XX
PN   WO9730158-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00290.
XX
PR   14-FEB-1996;   96FR-0001822.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425033/39.
DR   N-PSDB; AAT94550.
XX
PT   Recombinant protein containing the merozoite surface protein-1 p19
PT   fragment - useful in anti-malarial vaccines, diagnosis and protein
PT   purification
XX
PS   Disclosure; Fig 1B; 85pp; French.
XX
CC   This is the amino acid sequence of a recombinant protein comprising
CC   amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
CC   protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
CC   glycosylphosphatidylinositol membrane anchoring sequence.  p19 is the
CC   C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   116 AA;


  Query Match             30.0%;  Score 615;  DB 18;  Length 116;
  Best Local Similarity   96.5%;  Pred. No. 3.2e-36;
  Matches  110;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy     281 NISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADA 340
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       3 nisqhqcvkkqcpensgcfrhldereeckcllnykqegdkcvenpnptcnennggcdada 62

Qy     341 TCTEEDSGSSRKKITCECTKPDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           ||||||||: ||||||||||||||||||||||||||||||||||||||||||||
Db      63 kcteedsgsngkkitcectkpdsyplfdgifcsssnflgisfllilmlilysfi 116


RESULT  11
AAW22592
ID   AAW22592 standard; Protein; 116 AA.
XX
AC   AAW22592;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)A protein sequence.
XX
KW   Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          1..95
FT                   /note= "amino acids derived from P. falciparum MSP1 p19
FT                        fragment"
FT   Region          96..116
FT                   /note= "glycosylphosphatidylinositol anchoring sequence"
XX
PN   WO9730159-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00291.
XX
PR   14-FEB-1996;   96FR-0001821.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425034/39.
DR   P-PSDB; AAW22592.
XX
PT   Recombinant protein containing Plasmodium merozoite surface
PT   protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT   antibodies for diagnosis and protein purification
XX
PS   Disclosure; Fig 1B; 85pp; French.
XX
```

```
CC   This is the amino acid sequence of a recombinant protein comprising
CC   amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
CC   protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
CC   glycosylphosphatidylinositol membrane anchoring sequence. p19 is the
CC   C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   116 AA;


  Query Match           30.0%;  Score 615;  DB 18;  Length 116;
  Best Local Similarity   96.5%;  Pred. No. 3.2e-36;
  Matches  110;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    281 NISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADA 340
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      3 nisqhqcvkkqcpensgcfrhldereeckcllnykqegdkcvenpnptcnennggcdada 62

Qy    341 TCTEEDSGSSRKKITCECTKPDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          ||||||||: |||||||||||||||||||||||||||||||||||||||||||
Db     63 kcteedsgsngkkitcectkpdsyplfdgifcsssnflgisfllilmlilysfi 116


RESULT  12
AAB37609
ID   AAB37609 standard; Protein; 108 AA.
XX
AC   AAB37609;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Merozoite surface protein-119.
XX
KW   Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS   Plasmodium falciparum.
XX
PN   WO200063245-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-GB01558.
XX
PR   20-APR-1999;   99GB-0009072.
PR   13-MAY-1999;   99US-0311817.
PR   25-MAY-1999;   99CA-2271451.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR   WPI; 2001-015762/02.
DR   N-PSDB; AAC68977.
XX
PT   Novel variants of the C-terminal fragment of Plasmodium merozoite
PT   surface protein-1, useful as vaccines for treating or preventing
PT   malaria -
XX
PS   Example 5; Fig 15; 126pp; English.
XX
CC   The present invention relates to non-natural variants of a C-terminal
CC   fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC   non-natural variants have reduced affinity for at least 1 antibody
CC   capable of blocking a second antibody that inhibits the proteolytic
CC   cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC   one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC   MSP-1_4_2, compared to natural MSP-1_1_9. The non-natural variants of the
CC   present invention are useful for immunising a mammal against malaria, and
CC   can be used to treat malaria. The present sequence is MSP-119 protein.
XX
SQ   Sequence   108 AA;


  Query Match           25.8%;  Score 529;  DB 22;  Length 108;
  Best Local Similarity   88.6%;  Pred. No. 3.3e-30;
  Matches   93;  Conservative    4;  Mismatches    4;  Indels    4;  Gaps    1;

Qy    272 IEGKFQDMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCEENPNPTCNE 331
          |||:    ||:||||||||||:|||||||||||||||||||||||||||| |||||||||
Db      8 iegrh----niaqhqcvkkqcpqnsgcfrhldereeckcllnykqegdkcvenpnptcne 63

Qy    332 NNGGCDADATCTEEDSGSSRKKITCECTKPDSYPLFDGIFCSSSN 376
          ||||||||| ||||||||: |||||||||||||||||||||||||
Db     64 nnggcdadakcteedsgsngkkitcectkpdsyplfdgifcsssn 108


RESULT  13
AAB37608
ID   AAB37608 standard; protein; 96 AA.
XX
AC   AAB37608;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Merozoite surface protein-1.
XX
KW   Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS   Plasmodium falciparum.
XX
PN   WO200063245-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-GB01558.
XX
PR   20-APR-1999;   99GB-0009072.
PR   13-MAY-1999;   99US-0311817.
PR   25-MAY-1999;   99CA-2271451.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR   WPI; 2001-015762/02.
XX
PT   Novel variants of the C-terminal fragment of Plasmodium merozoite
PT   surface protein-1, useful as vaccines for treating or preventing
PT   malaria -
XX
PS   Example 2; Page 48; 126pp; English.
XX
CC   The present invention relates to non-natural variants of a C-terminal
CC   fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC   non-natural variants have reduced affinity for at least 1 antibody
CC   capable of blocking a second antibody that inhibits the proteolytic
CC   cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC   one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC   MSP-1_4_2, compared to natural MSP-1_1_9. The present sequence is the
CC   wild-type MSP-1 protein. This sequence was used to generate the variants
CC   of the present invention. The non-natural variants of the present
CC   invention are useful for immunising a mammal against malaria, and can be
CC   used to treat malaria.
XX
SQ   Sequence   96 AA;


  Query Match           25.7%;  Score 528;  DB 22;  Length 96;
  Best Local Similarity   94.8%;  Pred. No. 3.3e-30;
  Matches   91;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    281 NISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADA 340
          |||||||||||||:|||||||||||||||||||||||| ||||||||||||||||||
Db      1 nisqhqcvkkqcpqnsgcfrhldereeckcllnykqegdkcvenpnptcnennggcdada 60
```

```
Qy    341 TCTEEDSGSSRKKITCECTKPDSYPLFDGIFCSSSN 376
          ||||||||: |||||||||||||||||||||||||
Db     61 kcteedsgsngkkitcectkpdsyplfdgifcsssn 96


RESULT  14
AAW22593
ID   AAW22593 standard; Protein; 127 AA.
XX
AC   AAW22593;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)S protein sequence.
XX
KW   Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Chimeric - Plasmodium vivax.
OS   Chimeric - Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /note= "signal peptide"
FT   Protein         20..127
FT                   /note= "mature protein"
FT   Region          1..32
FT                   /note= "derived from P. vivax MSP1"
FT   Region          33..34
FT                   /note= "encoded by restriction enzyme sequence used to
FT                          create the chimeric sequence"
FT   Region          35..127
FT                   /note= "derived from P. falciparum C-terminal p19
FT                          fragment of MSP1"
XX
PN   WO9730159-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00291.
XX
PR   14-FEB-1996;   96FR-0001821.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425034/39.
DR   P-PSDB; AAW22592.
XX
PT   Recombinant protein containing Plasmodium merozoite surface
PT   protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT   antibodies for diagnosis and protein purification
XX
PS   Disclosure; Fig 1C; 85pp; French.
XX
CC   This is the amino acid sequence of a chimeric protein comprising amino
CC   acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
CC   linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC   C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC   p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   127 AA;

  Query Match             25.4%;  Score 521;  DB 18;  Length 127;
  Best Local Similarity   93.8%;  Pred. No. 1.5e-29;
  Matches   90;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    278 DMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCD 337
          |  ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     32 defnisqhqcvkkqcpensgcfrhldereeckcllnykqegdkcvenpnptcnennggcd 91

Qy    338 ADATCTEEDSGSSRKKITCECTKPDSYPLFDGIFCS 373
          ||| ||||||||: |||||||||||||||||||||
Db     92 adakcteedsgsngkkitcectkpdsyplfdgifcs 127


RESULT  15
AAW36102
ID   AAW36102 standard; Protein; 127 AA.
XX
AC   AAW36102;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)S protein sequence.
XX
KW   Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Chimeric - Plasmodium vivax.
OS   Chimeric - Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /note= "signal peptide"
FT   Protein         20..127
FT                   /note= "mature protein"
FT   Region          1..32
FT                   /note= "derived from P. vivax MSP1"
FT   Region          33..34
FT                   /note= "encoded by restriction enzyme sequence used to
FT                          create the chimeric sequence"
FT   Region          35..127
FT                   /note= "derived from P. falciparum C-terminal p19
FT                          fragment of MSP1"
XX
PN   WO9730158-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00290.
XX
PR   14-FEB-1996;   96FR-0001822.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425033/39.
DR   N-PSDB; AAT94549.
XX
PT   Recombinant protein containing the merozoite surface protein-1 p19
PT   fragment - useful in anti-malarial vaccines, diagnosis and protein
PT   purification
XX
PS   Disclosure; Fig 1C; 85pp; French.
XX
CC   This is the amino acid sequence of a chimeric protein comprising amino
CC   acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
CC   linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC   C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC   p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
```

SQ Sequence 127 AA;

Query Match 25.4%; Score 521; DB 18; Length 127;
Best Local Similarity 93.8%; Pred. No. 1.5e-29;
Matches 90; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```
Qy     278 DMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCD 337
           |  |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||  .
Db      32 defnisqhqcvkkqcpensgcfrhldereeckcllnykqegdkcvenpnptcnennggcd 91

Qy     338 ADATCTEEDSGSSRKKITCECTKPDSYPLFDGIFCS 373
           ||| |||||||: ||||||||||||||||||||||
Db      92 adakcteedsgsngkkitcectkpdsyplfdgifcs 127
```

Search completed: August 8, 2001, 12:31:55
Job time: 163 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:12 ; Search time 34.41 Seconds
(without alignments)
235.763 Million cell updates/sec

Title: US-09-500-376-2
Perfect score: 2062
Sequence: 1 AISVTMDNILSGFENEYDVI.....SNFLGISFLLILMLILYSFI 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	301	14.6	53	1	US-08-290-919-3	Sequence 3, Appli
2	282	13.7	53	1	US-08-290-919-4	Sequence 4, Appli
3	277	13.4	48	1	US-08-290-919-12	Sequence 12, Appl
4	274	13.3	48	1	US-08-290-919-2	Sequence 2, Appli
5	271	13.1	48	1	US-08-290-919-1	Sequence 1, Appli
6	266.5	12.9	106	1	US-08-290-919-11	Sequence 11, Appl
7	129.5	6.3	1494	3	US-08-755-587-186	Sequence 186, App
8	123.5	6.0	798	2	US-07-728-215-30	Sequence 30, Appl
9	123	6.0	740	1	US-08-257-073-5	Sequence 5, Appli
10	119	5.8	677	3	US-08-480-640A-115	Sequence 115, App
11	119	5.8	677	3	US-08-480-640A-193	Sequence 193, App
12	119	5.8	677	3	US-08-295-802-115	Sequence 115, App
13	119	5.8	677	4	US-08-686-968C-58	Sequence 58, Appl
14	119	5.8	677	4	US-08-686-968C-193	Sequence 193, App
15	119	5.8	677	4	US-08-488-237A-115	Sequence 115, App
16	119	5.8	677	4	US-08-488-237A-193	Sequence 193, App
17	118	5.7	1786	4	US-08-973-462-8	Sequence 8, Appli
18	117.5	5.7	414	5	PCT-US93-03077-3	Sequence 3, Appli
19	117.5	5.7	984	1	US-08-257-073-3	Sequence 3, Appli
20	117.5	5.7	984	2	US-08-184-009-120	Sequence 120, App
21	117.5	5.7	984	2	US-08-458-356-120	Sequence 120, App
22	117.5	5.7	1093	5	PCT-US93-03077-1	Sequence 1, Appli
23	116	5.6	540	1	US-08-286-325A-2	Sequence 2, Appli
24	116	5.6	3135	1	US-08-323-170B-2	Sequence 2, Appli
25	115.5	5.6	2710	1	US-08-480-604A-6	Sequence 6, Appli
26	115.5	5.6	2710	2	US-08-405-496A-6	Sequence 6, Appli
27	115	5.6	984	1	US-08-242-932-2	Sequence 2, Appli
28	115	5.6	984	1	US-08-714-481-2	Sequence 2, Appli
29	115	5.6	984	5	PCT-US95-06111-2	Sequence 2, Appli
30	114.5	5.6	1657	1	US-08-287-959-1	Sequence 1, Appli
31	111.5	5.4	1420	2	US-08-540-804-14	Sequence 14, Appl
32	111.5	5.4	1420	2	US-08-218-265-14	Sequence 14, Appl
33	111.5	5.4	1420	3	US-08-521-872-14	Sequence 14, Appl
34	111.5	5.4	1420	4	US-08-590-399-14	Sequence 14, Appl
35	111	5.4	976	4	US-09-104-324B-4	Sequence 4, Appli
36	111	5.4	1579	3	US-08-755-587-184	Sequence 184, App
37	111	5.4	3418	3	US-08-755-587-44	Sequence 44, Appl
38	110.5	5.4	1093	4	US-09-315-793-52	Sequence 52, Appl
39	109.5	5.3	1394	4	US-09-213-053-2	Sequence 2, Appli
40	108	5.2	2329	3	US-08-755-587-16	Sequence 16, Appl
41	108	5.2	3418	2	US-08-639-501-2	Sequence 2, Appli
42	108	5.2	3418	3	US-09-044-946-2	Sequence 2, Appli
43	108	5.2	3418	3	US-09-044-908-2	Sequence 2, Appli
44	107.5	5.2	2482	1	US-08-328-254-6	Sequence 6, Appli
45	107.5	5.2	3248	1	US-08-353-700-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-290-919-3
; Sequence 3, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Modified-site
```

```
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-3


  Query Match             14.6%;  Score 301;  DB 1;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-16;
  Matches   52;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    325 PNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSN 376
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 PNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSN 53


RESULT   2
US-08-290-919-4
; Sequence 4, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-4


  Query Match             13.7%;  Score 282;  DB 1;  Length 53;
  Best Local Similarity   94.2%;  Pred. No. 1.2e-14;
  Matches   49;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    325 PNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSN 376
          |||||||||||||||| |||||||| |||||||||||||||||||||||||
Db      2 PNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDSYPLFDGIFCSSSN 53


RESULT   3
US-08-290-919-12
; Sequence 12, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  48 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-290-919-12


  Query Match             13.4%;  Score 277;  DB 1;  Length 48;
  Best Local Similarity   97.9%;  Pred. No. 2.6e-14;
  Matches   47;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    281 NISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 328
          |||||||||||||:|||||||||||||||||||||||||||||||||||
Db      1 NISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 48


RESULT   4
US-08-290-919-2
; Sequence 2, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
```

```
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  48 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-2


  Query Match             13.3%;  Score 274;  DB 1;  Length 48;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-14;
  Matches   47;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     282 ISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 328
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 ISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 48


RESULT   5
US-08-290-919-1
; Sequence 1, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  48 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1
;      OTHER INFORMATION:  /label= X
;      OTHER INFORMATION:  /note= "X = M and N, or N"
US-08-290-919-1


  Query Match             13.1%;  Score 271;  DB 1;  Length 48;
  Best Local Similarity   97.9%;  Pred. No. 7.2e-14;
  Matches   46;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     282 ISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 328
           |||||||||||:|||||||||||||||||||||||||||||||||||
Db       2 ISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQEGDKCVENPNPT 48


RESULT   6
US-08-290-919-11
; Sequence 11, Application US/08290919
; Patent No. 5720959
;  GENERAL INFORMATION:
;    APPLICANT:  HOLDER, ANTHONY A.
;    APPLICANT:  BLACKMAN, MICHAEL J.
;    APPLICANT:  CHAPPEL, JONATHAN A.
;    TITLE OF INVENTION:  IMPROVEMENTS IN OR RELATING TO MALARIA
;    TITLE OF INVENTION:  VACCINE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,919
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9203821.5
;      FILING DATE:  22-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB93/00367
;      FILING DATE:  22-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  212242/HCM/MJL/6BC8/
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  106 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-290-919-11


  Query Match             12.9%;  Score 266.5;  DB 1;  Length 106;
  Best Local Similarity   49.0%;  Pred. No. 4.2e-13;
  Matches   50;  Conservative   17;  Mismatches   32;  Indels    3;  Gaps    3;

Qy    278 DMLNIS-QHQCV-KKQCPENSGCFRHLDEREECKCLLNYKQ-EGDKCVENPNPTCNENNG 334
          |:| :  :| ||  :  |:|:||||  :  || :||| ||: ||: |||| ||||: |||
Db      5 DLLGVDPKHVCVDTRDIPKNAGCFRDDNGTEEWRCLLGYKKGEGNTCVENNNPTCDINNG 64

Qy    335 GCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSN 376
          |||  | |   :|  | ||| | | :|     ::|:|||||:
Db     65 GCDPTASCQNAESTENSKKIICTCKEPTPNAYYEGVFCSSSS 106


RESULT   7
US-08-755-587-186
; Sequence 186, Application US/08755587
; Patent No. 6045997
;  GENERAL INFORMATION:
;    APPLICANT:  Futreal, Phillip A
;    APPLICANT:  Wooster, Richard F
;    APPLICANT:  Ashworth, Alan
;    APPLICANT:  Stratton, Michael R
;    TITLE OF INVENTION:  Materials and methods relating to the
;    TITLE OF INVENTION:  identification and sequencing of the BRCA2 cancer
;    TITLE OF INVENTION:  susceptibility gene and uses thereof.
;    NUMBER OF SEQUENCES:  222
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson
;      STREET:  310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
;      CITY:  Raleigh
;      STATE:  NC
;      COUNTRY:  USA
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/755,587
;      FILING DATE:  25-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9523959.6
;      FILING DATE:  23-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9525555.0
;      FILING DATE:  14-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9617961.9
;      FILING DATE:  28-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kenneth D Sibley
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5405-135
;  INFORMATION FOR SEQ ID NO:  186:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1494 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-755-587-186


  Query Match             6.3%;  Score 129.5;  DB 3;  Length 1494;
  Best Local Similarity   22.9%;  Pred. No. 0.17;
  Matches   95;  Conservative   68;  Mismatches  167;  Indels   85;  Gaps   22;

Qy      4 VTMDNILSGFEN------EYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLN---DILNSR 54
          ||| |      |      ::|  ||       | :    |||     :||    :|| :
Db    760 VTMSNKQQLTANTGQNIKDFDTFYL--------SFQTASRKNIRVSRESLNKARSLLNQK 811

Qy     55 LKKRKY--FLDVLESDLMQFKHIS----SNEYIIEDSFKLLNSEQKNTLLKSYKYIKESV 108
            : :   | | | |:|:    |     ||  :||      |:|:|: : |     |::
Db    812 WTEEELNNFSDSLNSELLPGIDIKKTDISNHEVIE------NTERKDKITKESLIGTENI 865

Qy    109 --------ENDIKFAQEG--ISYYE---KVLAKYKDDLESIKKVIKEEKEKFPSSPPTTP 155
                  |: ||  :|   : ::    | :   |: |: :| :  ||||:  |
Db    866 LLILQQRPESKIKKIKESAVLGFHTASGKKIEITKESLDKVKNLF-EEKEQDNSEITNFS 924

Qy    156 PSPAKTDEQKKESK----------FLPFLTNIETLYNNL-VNKIDDYLINLKAKINDCNV 204
            ||  : ::| |            :      |  :::|   |: : :  |: ::   |:
Db    925 HRGAKMSKDREECKDGRELACGTTEITTTPEYEETHSSLEKKKVSNEIAALRPRLLSDNL 984

Qy    205 EKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDDTKKDMLGKLLSTGLVQNFP 264
          |   ::||:  :       |:|: :|  : |  |       |  |     :  ::| |
Db    985 YKQTENLKISDHA-----SQKVDVHEN-TEKETAK-------KPTMYTNQSTYSAIENSP 1031

Qy    265 NTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVEN 324
          |  ::  | ||    ::|:   : :     | :    ||      :| |:  |  ||
Db   1032 LT-FTQDTEEKF----SVSEASLFEAKKWLREGEWDDQSERINA-AKVNCKEYPDDYVE- 1084

Qy    325 PNPTCNENNGGCDADAKCTEED----SGSNGKKITCECTKPDSYPLFDGIFCSSS 375
          ||:|     |  :::  || |     |   |       |  :||    | || ||
Db   1085 -NPSC-----GNSSNSAITENDKNHLSEKQGSTYLSNSTMSNSYSYHPG-FCHSS 1132


RESULT   8
US-07-728-215-30
; Sequence 30, Application US/07728215
; Patent No. 5962643
;  GENERAL INFORMATION:
;    APPLICANT:  Sheppard, Dean
;    APPLICANT:  Quaranta, Vito
;    APPLICANT:  Pytela, Robert
;    TITLE OF INVENTION:  A No. 5962643el Integrin Beta Subunit and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/728,215
;      FILING DATE:  19910711
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P31 8717
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  798 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-728-215-30


  Query Match             6.0%;  Score 123.5;  DB 2;  Length 798;
  Best Local Similarity   20.8%;  Pred. No. 0.21;
  Matches  103;  Conservative   67;  Mismatches  146;  Indels  179;  Gaps   29;

Qy      48 NDILNSR----LKKRKYFLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKY 103
           :|| | |    :|| |   :  :    : |    ::   :     | | :  |    :|
Db      78 DDIENPRGSKDIKKNKNVTNRSKGTAEKLKPEDIHQIQPQQLVLRLRSGEPQTFTLKFKR 137

Qy     104 IKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKV---IKEEKEKFPSS---------- 150
             |    |: :  : :||        ||||::| :   :  |  :  |
Db     138 -AEDYPIDLYYLMD-LSY------SMKDDLENVKSLGTDLMNEMRRITSDFRIGFGSFVE 189

Qy     151 ---PPTTPPSPAK-----TDEQKKESKF----LPFLTNIETLYNNLVNK------ID--- 189
               |    :|||     | ||   : |    :  |||   ::| || |      :|
Db     190 KTVMPYISTTPAKLRNPCTSEQNCTTPFSYKNVLSLTNKGEVFNELVGKQRISGNLDSPE 249

Qy     190 ---DYLINL----------------------------KAKI------ND--CNVEKDE-- 208
             | :: :                             |:      ||  |::| :
Db     250 GGFDAIMQVAVCGSLIGWRNVTRLLVFSTDAGFHFAGDGKLGGIVLPNDGQCHLENNMYT 309

Qy     209 ----------AHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDDTKKDMLGKLLSTG 258
                    ||: : |||:    ::   :|    :|: : | : :   |  :| | :
Db     310 MSHYYDYPSIAHL-VQKLSE----NNIQTIFAVTEEFQPVYKELKNLIPKSAVGTLSANS 364

Qy     259 -----LVQNFPNTIISKLI--EGKFQDMLNISQHQCVK---------------------- 289
                |: :  |:: |::|   ||  : : ||     |
Db     365 SNVIQLIIDAYNSLSSEVILENGKLSEGVTISYKSYCKNGVNGTGENGRKCSNISIGDEV 424

Qy     290 --------KQCP-ENSGCFR-----HLDERE-------ECKCLLNYKQEGDKCVENPNPT 328
                  :|| ::|  |:       :| |        ||:|      |  || |  | |
Db     425 QFEISITSNKCPKKDSDSFKIRPLGFTEEVEVILQYICECECQSEGIPESPKCHEG-NGT 483

Qy     329 -------CNENNGG----C--------DADAKCTEEDS----GSNGKKI--TCECTKPD- 362
                 |||   |    |         | || | :|:|     :||: :   | | | |
Db     484 FECGACRCNEGRVGRHCECSTDEVNSEDMDAYCRKENSSEICSNNGECVCGQCVCRKRDN 543

Qy     363 SYPLFDGIFCSSSNF 377
           :  :: | ||   ||
Db     544 TNEIYSGKFCECDNF 558


RESULT   9
US-08-257-073-5
; Sequence 5, Application US/08257073
; Patent No. 5766597
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  de Taisne, Charles
;    APPLICANT:  Tine, John A.
;    TITLE OF INVENTION:  MALARIA RECOMBINANT POXVIRUS VACCINE
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue, 25th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/257,073
;      FILING DATE:  09-JUN-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/075,783
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/852,305
;      FILING DATE:  18-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,183
;      FILING DATE:  20-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2570
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066 CURTMS
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  740 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-257-073-5


  Query Match             6.0%;  Score 123;  DB 1;  Length 740;
  Best Local Similarity   22.9%;  Pred. No. 0.21;
  Matches   96;  Conservative   53;  Mismatches  110;  Indels  160;  Gaps   22;

Qy      47 LNDILNSRLKKR---KYFLDVLESDLMQFKHISSNEYI------IEDSFKLLNSEQ---- 93
           :||:||  :|       |:||::| | |   : ||: |·      |    |||  :
Db     289 VNDVLNVNMKNMITGNNFMDVVKSTLAQSGGLGSNDLINFLNQGKEIGENLLNITKMNLG 348

Qy      94 -KNTL-------------LKSYKYI----KESVENDIK------FAQEGISY-----YE 123
            || |              | :|::|    | || | :|        :  :||     |
Db     349 DKNNLESFPLDQLNMLKDNLINYEFILNNLKTSVLNKLKDLLLRLLYKAYVSYKKRKAQE 408

Qy     124 KVL-------AKYKDDLE------------SIKKVIKEEKEK-FPSSPPTTPPSPAKTD 162
           | |        :| ::|:                :| ::|: | | ||           |:
Db     409 KGLPEPTVTNEEYVEELKKGILDMGIKLLFSKVKSLLKKLKNKIFPKKKEDNQAVDTKSM 468

Qy     163 EQKKESKFLPFL--------TNIETLYNNLVNKIDDYLINLKAKINDCNV---------- 204
           |: |  |  | |         :||    ||::::|| :   |    |  ||
Db     469 EEPK-VKAQPALRGVEPTEDSNIMNSINNVMDEIDFFEKELIENNNTPNVVPPTQSKKKN 527

Qy     205 --------------------------EKDEAHVKITKLS-DLKAIDDKIDLFKNHNDFE 236
                                       | |:  ||: |:   ||    | :  || |  |
Db     528 KNETVSGMDENFDNHPENYFKEEYYYDENDDMEVKVKKIGVTLK----KFEPLKNGNVSE 583

Qy     237 AIK-----------KLINDDTK--KDMLGKLLSTGLVQNFPNTIISKLIEGKFQ----- 277
           ||            : ||:| :  |  |  : :   : |: |   :| |:  ||
Db     584 TIKLIHLGNKDKKHIEAINNDIQIIKQELQAIYNE--LMNYTNG--NKNIQQIFQQNILE 639

Qy     278 -DMLNISQHQCVKKQ--------------CPENSGCFRHLDEREECKCLLNYKQEGDK 320
            |:||    : ::||               |:|       :| |: |     |:| :|
Db     640 NDVLNQETEEEMEKQVEAITKQIEAEVDALAPKNK------EEEEKEKEKEKEKEEKEK 692
```

```
RESULT 10
US-08-480-640A-115
; Sequence 115, Application US/08480640A
; Patent No. 6033904
;  GENERAL INFORMATION:
;    APPLICANT:  Cochran, Mark D.
;    APPLICANT:  Junker, David E.
;    TITLE OF INVENTION:  Recombinant Swinepox Virus
;    NUMBER OF SEQUENCES:  225
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,640A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  115:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  677 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Swinepox virus
;      STRAIN:  Kasza
;      INDIVIDUAL ISOLATE:  S-SPV-001
;    IMMEDIATE SOURCE:
;      CLONE:  515-85.1
;    POSITION IN GENOME:
;      MAP POSITION:  -23.2
;      UNITS:  %G
US-08-480-640A-115


  Query Match             5.8%;  Score 119;  DB 3;  Length 677;
  Best Local Similarity   22.2%;  Pred. No. 0.37;
  Matches   76;  Conservative   65;  Mismatches  137;  Indels   64;  Gaps   18;

Qy      18 DVIYLKPLAGVYRSLKKQIEKNIFTF----NLNLNDILNSRLKKRKYFLDVLE---SDLM 70
           |  :: |     ||| : | : |       | |: | ::  :|    |: :|    |::
Db     185 DAFHIPPTHISLRSL-RDINRIIELLKKYPNNNIIDYISDSIKSNSSFIHILHMIISNMF 243

Qy      71 QFKHISSNEYI--IEDSFKLLNSEQ-KNTLLKSYKY----IKESVENDIKFAQEGISYYE 123
             | |::|  : |  :|:|    |   : ||      ::   ::|  | :: || |:
Db     244 PAIIPSVNDFISTVVDKDRLINMYGIKCVAMFSYDINMIDLESLDDSDYIFIEKNISIYD 303

Qy     124 KVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNN 183
             | :|    |: :| || :  ::         | ::  :  |        :   |:
Db     304 ---VKCRDFANMIRDKVKREKNRILTT---------KCEDIIRYIKLFS-----KNRIND 346

Qy     184 LVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLIN 243
           ||::: ||::       || |:        |||: :: |:  |    |        |::
Db     347 ENNKVEEVLIHID------NVSKNNK----LSLSDISSLMDQFRL----NPCTIRNILLS 392

Qy     244 DDTKKDMLGKLLSTGLVQNF-----PNTIISKLIEGKF-QDMLNISQHQCVK--KQCPEN 295
             | |    |||:   |:|:      |  : | |:|    ||::      :|  ||  :
Db     393 SATIK---SKLLALRAVKNWKCYSLTNVSMYKKIKGVIVMDMVDYISTNILKYHKQLYDK 449

Qy     296 SGCFRHLDEREECKC------LLNYKQEGDKCVENPNPTCNE 331
             | :  : : |||      : ::  |   |: |   | |:
Db     450 MSTFEYKRDIKSCKCSICSDSITHHIYETTSCI-NYKSTDND 490


RESULT 11
US-08-480-640A-193
; Sequence 193, Application US/08480640A
; Patent No. 6033904
;  GENERAL INFORMATION:
;    APPLICANT:  Cochran, Mark D.
;    APPLICANT:  Junker, David E.
;    TITLE OF INVENTION:  Recombinant Swinepox Virus
;    NUMBER OF SEQUENCES:  225
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,640A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  193:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  677 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-640A-193


  Query Match             5.8%;  Score 119;  DB 3;  Length 677;
  Best Local Similarity   22.2%;  Pred. No. 0.37;
  Matches   76;  Conservative   65;  Mismatches  137;  Indels   64;  Gaps   18;

Qy      18 DVIYLKPLAGVYRSLKKQIEKNIFTF----NLNLNDILNSRLKKRKYFLDVLE---SDLM 70
           |  :: |     ||| : | : |       | |: | ::  :|    |: :|    |::
Db     185 DAFHIPPTHISLRSL-RDINRIIELLKKYPNNNIIDYISDSIKSNSSFIHILHMIISNMF 243

Qy      71 QFKHISSNEYI--IEDSFKLLNSEQ-KNTLLKSYKY----IKESVENDIKFAQEGISYYE 123
             | |::|  : |  :|:|    |   : ||      ::   ::|  | :: || |:
Db     244 PAIIPSVNDFISTVVDKDRLINMYGIKCVAMFSYDINMIDLESLDDSDYIFIEKNISIYD 303

Qy     124 KVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNN 183
             | :|    |: :| || :  ::         | ::  :  |        :   |:
Db     304 ---VKCRDFANMIRDKVKREKNRILTT---------KCEDIIRYIKLFS-----KNRIND 346

Qy     184 LVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLIN 243
           ||::: ||::       || |:        |||: :: |:  |    |        |::
Db     347 ENNKVEEVLIHID------NVSKNNK----LSLSDISSLMDQFRL----NPCTIRNILLS 392

Qy     244 DDTKKDMLGKLLSTGLVQNF-----PNTIISKLIEGKF-QDMLNISQHQCVK--KQCPEN 295
             | |    |||:   |:|:      |  : | |:|    ||::      :|  ||  :
Db     393 SATIK---SKLLALRAVKNWKCYSLTNVSMYKKIKGVIVMDMVDYISTNILKYHKQLYDK 449
```

```
Qy    296 SGCFRHLDEREECKC------LLNYKQEGDKCVENPNPTCNE 331
          | :  : : |||      : ::  |   |: |   | |:
Db    450 MSTFEYKRDIKSCKCSICSDSITHHIYETTSCI-NYKSTDND 490


RESULT 12
US-08-295-802-115
; Sequence 115, Application US/08295802
; Patent No. 6127163
;  GENERAL INFORMATION:
;    APPLICANT:  Cochran Ph.D., Mark D
;    APPLICANT:  Junker M.S., David E
;    TITLE OF INVENTION:  Recombinant Swinepox Virus
;    NUMBER OF SEQUENCES:  188
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/295,802
;      FILING DATE:  Herewith
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)977-9550
;      TELEFAX:  (212)664-0525
;      TELEX:  422523
;  INFORMATION FOR SEQ ID NO:  115:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  677 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Swinepox virus
;      STRAIN:  Kasza
;      INDIVIDUAL ISOLATE:  S-SPV-001
;    IMMEDIATE SOURCE:
;      CLONE:  515-85.1
;    POSITION IN GENOME:
;      MAP POSITION:  ~23.2
;      UNITS:  %G
US-08-295-802-115


  Query Match             5.8%;  Score 119;  DB 3;  Length 677;
  Best Local Similarity   22.2%;  Pred. No. 0.37;
  Matches   76;  Conservative   65;  Mismatches  137;  Indels   64;  Gaps   18;

Qy     18 DVIYLKPLAGVYRSLKKQIEKNIFTF----NLNLNDILNSRLKKRKYFLDVLE---SDLM 70
          |  :: |     ||| : | : |       | |: | ::  :|    |: :|    |::
Db    185 DAFHIPPTHISLRSL-RDINRIIELLKKYPNNNIIDYISDSIKSNSSFIHILHMIISNMF 243

Qy     71 QFKHISSNEYI--IEDSFKLLNSEQ-KNTLLKSYKY----IKESVENDIKFAQEGISYYE 123
               | |::|  : |  :|:|    |   : ||      ::   ::|  | :: || |:
Db    244 PAIIPSVNDFISTVVDKDRLINMYGIKCVAMFSYDINMIDLESLDDSDYIFIEKNISIYD 303

Qy    124 KVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNN 183
            | :|    |:  :| || :  ::       | ::  :  |         :   |:
Db    304 ---VKCRDFANMIRDKVKREKNRILTT---------KCEDIIRYIKLFS-----KNRIND 346

Qy    184 LVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLIN 243
            ||::: ||::      || |:          |||: :: |:  |    |       |::
Db    347 ENNKVEEVLIHID------NVSKNNK----LSLSDISSLMDQFRL----NPCTIRNILLS 392

Qy    244 DDTKKDMLGKLLSTGLVQNF-----PNTIISKLIEGKF-QDMLNISQHQCVK--KQCPEN 295
            | |    |||:   |:|:      |  : | |:|    ||::      :|  ||  :
Db    393 SATIK---SKLLALRAVKNWKCYSLTNVSMYKKIKGVIVMDMVDYISTNILKYHKQLYDK 449

Qy    296 SGCFRHLDEREECKC------LLNYKQEGDKCVENPNPTCNE 331
          | :  : : |||      : ::  |   |: |   | |:
Db    450 MSTFEYKRDIKSCKCSICSDSITHHIYETTSCI-NYKSTDND 490


RESULT 13
US-08-686-968C-58
; Sequence 58, Application US/08686968C
; Patent No. 6221361
; GENERAL INFORMATION:
;  APPLICANT: Cochran, Mark D.
;  APPLICANT:  Junker, David E.
;  TITLE OF INVENTION: Recombinant Swinepox Virus
;  FILE REFERENCE: 39119-H/JML
;  CURRENT APPLICATION NUMBER: US/08/686,968C
;  CURRENT FILING DATE:  1996-07-25
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
;   LENGTH: 677
;   TYPE: PRT
;   ORGANISM: Swinepox virus
US-08-686-968C-58


  Query Match             5.8%;  Score 119;  DB 4;  Length 677;
  Best Local Similarity   22.2%;  Pred. No. 0.37;
  Matches   76;  Conservative   65;  Mismatches  137;  Indels   64;  Gaps   18;

Qy     18 DVIYLKPLAGVYRSLKKQIEKNIFTF----NLNLNDILNSRLKKRKYFLDVLE---SDLM 70
          |  :: |     ||| : | : |       | |: | ::  :|    |: :|    |::
Db    185 DAFHIPPTHISLRSL-RDINRIIELLKKYPNNNIIDYISDSIKSNSSFIHILHMIISNMF 243

Qy     71 QFKHISSNEYI--IEDSFKLLNSEQ-KNTLLKSYKY----IKESVENDIKFAQEGISYYE 123
               | |::|  : |  :|:|    |   : ||      ::   ::|  | :: || |:
Db    244 PAIIPSVNDFISTVVDKDRLINMYGIKCVAMFSYDINMIDLESLDDSDYIFIEKNISIYD 303

Qy    124 KVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNN 183
            | :|    |:  :| || :  ::       | ::  :  |         :   |:
Db    304 ---VKCRDFANMIRDKVKREKNRILTT---------KCEDIIRYIKLFS-----KNRIND 346

Qy    184 LVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLIN 243
            ||::: ||::      || |:          |||: :: |:  |    |       |::
Db    347 ENNKVEEVLIHID------NVSKNNK----LSLSDISSLMDQFRL----NPCTIRNILLS 392

Qy    244 DDTKKDMLGKLLSTGLVQNF-----PNTIISKLIEGKF-QDMLNISQHQCVK--KQCPEN 295
            | |    |||:   |:|:      |  : | |:|    ||::      :|  ||  :
Db    393 SATIK---SKLLALRAVKNWKCYSLTNVSMYKKIKGVIVMDMVDYISTNILKYHKQLYDK 449

Qy    296 SGCFRHLDEREECKC------LLNYKQEGDKCVENPNPTCNE 331
          | :  : : |||      : ::  |   |: |   | |:
Db    450 MSTFEYKRDIKSCKCSICSDSITHHIYETTSCI-NYKSTDND 490


RESULT 14
US-08-686-968C-193
; Sequence 193, Application US/08686968C
; Patent No. 6221361
; GENERAL INFORMATION:
;  APPLICANT: Cochran, Mark D.
;  APPLICANT:  Junker, David E.
;  TITLE OF INVENTION: Recombinant Swinepox Virus
;  FILE REFERENCE: 39119-H/JML
;  CURRENT APPLICATION NUMBER: US/08/686,968C
```

```
;  CURRENT FILING DATE:  1996-07-25
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 193
;   LENGTH: 677
;   TYPE: PRT
;   ORGANISM: Swinepox virus
US-08-686-968C-193


  Query Match             5.8%;  Score 119;  DB 4;  Length 677;
  Best Local Similarity   22.2%;  Pred. No. 0.37;
  Matches   76;  Conservative   65;  Mismatches  137;  Indels   64;  Gaps   18;

Qy      18 DVIYLKPLAGVYRSLKKQIEKNIFTF----NLNLNDILNSRLKKRKYFLDVLE---SDLM 70
           |  :: |     ||| : | : |       | |: | ::  :|    |: :|    |::
Db     185 DAFHIPPTHISLRSL-RDINRIIELLKKYPNNNIIDYISDSIKSNSSFIHILHMIISNMF 243

Qy      71 QFKHISSNEYI--IEDSFKLLNSEQ-KNTLLKSYKY----IKESVENDIKFAQEGISYYE 123
             | |::|  : |  :|:|    |   : ||      .::   ::|  | :: || |:
Db     244 PAIIPSVNDFISTVVDKDRLINMYGIKCVAMFSYDINMIDLESLDDSDYIFIEKNISIYD 303

Qy     124 KVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNN 183
             | :|    |:  :| || :  ::         | ::  :  |        :   |:
Db     304 ---VKCRDFANMIRDKVKREKNRILTT---------KCEDIIRYIKLFS-----KNRIND 346

Qy     184 LVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLIN 243
           ||::: ||::       || |:        |||: :: |:  |    |       |::
Db     347 ENNKVEEVLIHID------NVSKNNK----LSLSDISSLMDQFRL----NPCTIRNILLS 392

Qy     244 DDTKKDMLGKLLSTGLVQNF-----PNTIISKLIEGKF-QDMLNISQHQCVK--KQCPEN 295
           | |    |||:   |:|:      |  : | |:|    ||::      :|  ||  :
Db     393 SATIK---SKLLALRAVKNWKCYSLTNVSMYKKIKGVIVMDMVDYISTNILKYHKQLYDK 449

Qy     296 SGCFRHLDEREECKC------LLNYKQEGDKCVENPNPTCNE 331
           | :  : : |||      : ::  |  |: |   | |:
Db     450 MSTFEYKRDIKSCKCSICSDSITHHIYETTSCI-NYKSTDND 490


RESULT  15
US-08-488-237A-115
; Sequence 115, Application US/08488237A
; Patent No. 6251403
;  GENERAL INFORMATION:
;    APPLICANT:  Cochran, Mark D.
;    APPLICANT:  Junker, David E.
;    TITLE OF INVENTION:  Recombinant Swinepox Virus
;    NUMBER OF SEQUENCES:  225
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,237A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  115:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  677 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Swinepox virus
;      STRAIN:  Kasza
;      INDIVIDUAL ISOLATE:  S-SPV-001
;    IMMEDIATE SOURCE:
;      CLONE:  515-85.1
;    POSITION IN GENOME:
;      MAP POSITION:  ~23.2
;      UNITS:  %G
US-08-488-237A-115


  Query Match             5.8%;  Score 119;  DB 4;  Length 677;
  Best Local Similarity   22.2%;  Pred. No. 0.37;
  Matches   76;  Conservative   65;  Mismatches  137;  Indels   64;  Gaps   18;

Qy      18 DVIYLKPLAGVYRSLKKQIEKNIFTF----NLNLNDILNSRLKKRKYFLDVLE---SDLM 70
           |  :: |     ||| : | : |       | |: | ::  :|    |: :|    |::
Db     185 DAFHIPPTHISLRSL-RDINRIIELLKKYPNNNIIDYISDSIKSNSSFIHILHMIISNMF 243

Qy      71 QFKHISSNEYI--IEDSFKLLNSEQ-KNTLLKSYKY----IKESVENDIKFAQEGISYYE 123
             | |::|  : |  :|:|    |   : ||      ::   ::|  | :: || |:
Db     244 PAIIPSVNDFISTVVDKDRLINMYGIKCVAMFSYDINMIDLESLDDSDYIFIEKNISIYD 303

Qy     124 KVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNN 183
             | :|    |:  :| || :  ::         | ::  :  |        :   |:
Db     304 ---VKCRDFANMIRDKVKREKNRILTT---------KCEDIIRYIKLFS-----KNRIND 346

Qy     184 LVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLIN 243
           ||::: ||::       || |:        |||: :: |:  |    |       |::
Db     347 ENNKVEEVLIHID------NVSKNNK----LSLSDISSLMDQFRL----NPCTIRNILLS 392

Qy     244 DDTKKDMLGKLLSTGLVQNF-----PNTIISKLIEGKF-QDMLNISQHQCVK--KQCPEN 295
           | |    |||:   |:|:      |  : | |:|    ||::      :|  ||  :
Db     393 SATIK---SKLLALRAVKNWKCYSLTNVSMYKKIKGVIVMDMVDYISTNILKYHKQLYDK 449

Qy     296 SGCFRHLDEREECKC------LLNYKQEGDKCVENPNPTCNE 331
           | :  : : |||      : ::  |  |: |   | |:
Db     450 MSTFEYKRDIKSCKCSICSDSITHHIYETTSCI-NYKSTDND 490


Search completed: August  8, 2001, 12:29:56
Job time: 44 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:12 ; Search time 60.73 Seconds
(without alignments)
393.312 Million cell updates/sec

Title: US-09-500-376-2
Perfect score: 2062
Sequence: 1 AISVTMDNILSGFENEYDVI.....SNFLGISFLLILMLILYSFI 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1283.5	62.2	1639	19	AAW54145	P. falciparum synt
2	1191.5	57.8	355	20	AAY09372	Merozoite surface
3	1191.5	57.8	355	20	AAY05832	Merozoite surface
4	1191.5	57.8	361	20	AAY09373	Merozoite surface
5	1191.5	57.8	361	20	AAY05833	Merozoite surface
6	1184.5	57.4	376	20	AAY09374	Modified merozoite
7	1184.5	57.4	376	20	AAY05834	Modified merozoite
8	1124	54.5	1654	6	AAP50777	Sequence of the P1
9	651.5	31.6	262	22	AAB37610	Merozoite surface
10	640	31.0	116	18	AAW36103	PfMSP1(p19)A prote
11	640	31.0	116	18	AAW22592	PfMSP1(p19)A prote
12	554	26.9	108	22	AAB37609	Merozoite surface
13	553	26.8	96	22	AAB37608	Merozoite surface
14	546	26.5	127	18	AAW22593	PfMSP1(p19)S prote
15	546	26.5	127	18	AAW36102	PfMSP1(p19)S prote
16	307	14.9	54	14	AAR41356	MSP1EGF2A EGF2-lik
17	288	14.0	54	14	AAR41357	MSP1EGF2B EGF2-lik
18	282	13.7	49	14	AAR41355	MSP1EGF1B EGF1-lik
19	279	13.5	49	14	AAR41354	MSP1EGF1A EGF1-lik
20	279	13.5	93	22	AAB37611	Merozoite surface
21	266.5	12.9	106	14	AAR41358	P. yoelii combined
22	200.5	9.7	350	21	AAY70278	Recombinant vaccin
23	180.5	8.8	980	21	AAB18294	Plasmodium falcipa
24	138.5	6.7	1979	21	AAB18171	Plasmodium falcipa
25	136	6.6	2573	21	AAB18234	Plasmodium falcipa
26	133	6.5	1712	21	AAB18205	Plasmodium falcipa
27	130	6.3	2500	21	AAB18272	Plasmodium falcipa
28	129.5	6.3	1817	21	AAB18255	Plasmodium falcipa
29	128	6.2	1254	11	AAR07503	Merozite apical-en
30	128	6.2	1254	18	AAW24575	Merozoite apical-e
31	127.5	6.2	1997	21	AAB18287	Plasmodium falcipa
32	127	6.2	26	22	AAB71974	Plasmodium falcipa
33	127	6.2	1166	21	AAB18268	Plasmodium falcipa
34	125.5	6.1	2485	21	AAB18172	Plasmodium falcipa
35	125	6.1	696	21	AAB18181	Plasmodium falcipa
36	125	6.1	2206	21	AAB18254	Plasmodium falcipa
37	125	6.1	2285	20	AAW98149	Bacillus subtilis
38	124	6.0	558	21	AAB18273	Plasmodium falcipa
39	123.5	6.0	798	22	AAB36937	Human integrin bet
40	123.5	6.0	963	19	AAW70540	Integrin beta-1 ch
41	123.5	6.0	1188	21	AAB18183	Plasmodium falcipa
42	123	6.0	400	18	AAW55691	H. pylori ORF 06gp
43	123	6.0	407	21	AAB18248	Plasmodium falcipa
44	123	6.0	740	13	AAR27530	Plasmodium falcipa
45	123	6.0	740	16	AAR68838	Plasmodium falcipa

ALIGNMENTS

RESULT 1
AAW54145
ID AAW54145 standard; Protein; 1639 AA.
XX
AC AAW54145;
XX
DT 23-SEP-1998 (first entry)
XX
DE P. falciparum synthetic gp190 protein.
XX
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW monoclonal antibody; passive immunisation; parasite.
XX
OS Plasmodium falciparum.
OS Synthetic.
XX
PN WO9814583-A2.
XX
PD 09-APR-1998.
XX
PF 02-OCT-1997; 97WO-EP05441.
XX
PR 02-OCT-1996; 96DE-4040817.
XX
PA (BUJA/) BUJARD H.
XX
PI Bujard H, Pan W, Tolle R;
XX
DR WPI; 1998-240088/21.
DR N-PSDB; AAV21451, AAV35363.
XX
PT Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PT reducing their AT content

XX
PS Example 1; Fig 3c; 48pp; German.
XX
CC This sequence represents a modified Plasmodium falciparum gp190/MSP-1
CC (merozoite surface) protein. The gene encoding this protein has been
CC stabilised by reducing the AT content of the nucleotide sequence. Such a
CC protein is useful in vaccines against malaria or for producing monoclonal
CC antibodies (for passive immunisation). The complete gp190 protein can now
CC be produced outside the parasite and has, at least over extended regions,
CC the native pattern of folding. Larger amounts of the protein can be
CC produced recombinantly than would be possible using the parasites as
CC source.
XX
SQ Sequence 1639 AA;

Query Match 62.2%; Score 1283.5; DB 19; Length 1639;
Best Local Similarity 62.1%; Pred. No. 5.7e-82;
Matches 242; Conservative 54; Mismatches 73; Indels 21; Gaps 2;

Qy 6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
Db 1270 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 1329

Qy 66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
Db 1330 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 1389

Qy 126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
Db 1390 sekyksdldsikkyi----------ndkqgenekylpflnnietlyktvn 1429

Qy 186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
Db 1430 dkidlfvihleakvlnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 1489

Qy 246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDE 304
Db 1490 nhnnlltkflstgmvfenlaktvlsnlldgnlqgmlnisqhqcvkkqcpqnsgcfrhlde 1549

Qy 305 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 364
Db 1550 reeckcllnykqegdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsy 1609

Qy 365 PLFDGIFCSSSNFLGISFLLILMLILYSFI 394
Db 1610 plfdgifcsssnflgiffllilmlilysfi 1639

RESULT 2
AAY09372
ID AAY09372 standard; Protein; 355 AA.
XX
AC AAY09372;
XX
DT 31-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX
OS Plasmodium falciparum.
XX
PN WO9920774-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22226.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI; 1999-288313/24.
DR P-PSDB; AAX56008.
XX
PT Modified malarial protein for use in anti-malarial vaccines
XX
PS Example; Fig 1; 35pp; English.
XX
CC The present sequence represents a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
CC has been modified (see AAX56008) compared to the native sequence (see
CC AAX56009) such that 306 nucleotide positions have been replaced to
CC lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
CC instability motifs while maintaining the same protein amino acid
CC sequence. These alterations allow MSP-1-42 to be expressed in
CC mammalian cell culture and in transgenic mice. Native MSP-1-12
CC is known to be difficult to express in cell culture systems,
CC mammalian cell culture systems or in transgenic animals. The
CC invention allows expression of MSP-1 protein in the milk of
CC transgenic animals, and also provides a DNA vaccine comprising a
CC vector containing the altered MSP-1-42 sequence.
XX
SQ Sequence 355 AA;

Query Match 57.8%; Score 1191.5; DB 20; Length 355;
Best Local Similarity 60.2%; Pred. No. 2.4e-76;
Matches 222; Conservative 54; Mismatches 72; Indels 21; Gaps 2;

Qy 6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
Db 7 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 66

Qy 66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
Db 67 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 126

Qy 126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
Db 127 sekyksdldsikkyi----------ndkqgenekylpflnnietlyktvn 166

Qy 186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
Db 167 dkidlfvihleakvlnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 226

Qy 246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDE 304
Db 227 nhnnlltkflstgmvfenlaktvlsnlldgnlqgmlnisqhqcvkkqcpqnsgcfrhlde 286

Qy 305 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 364
Db 287 reeckcllnykqegdkcvenpnptcnenngcdadakcteedsgsngkkitcectkpdsy 346

Qy 365 PLFDGIFCS 373
Db 347 plfdgifcs 355

RESULT 3
AAY05832
ID AAY05832 standard; Protein; 355 AA.
XX
AC AAY05832;
XX
DT 02-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42.

```
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal.
XX
OS   Plasmodium falciparum.
XX
PN   WO9920766-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22225.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-302742/25.
DR   N-PSDB; AAX25586.
XX
PT   New modified recombinant nucleic acid sequences useful for producing
PT   malarial DNA vaccine
XX
PS   Disclosure; Fig 1; 43pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The nucleic acid sequence encoding MSP-1-42
CC   has been modified (see AAX25586) compared to the native sequence (see
CC   AAX25587) such that 306 nucleotide positions have been replaced to
CC   lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
CC   instability motifs while maintaining the same protein amino acid
CC   sequence.  These alterations allow MSP-1-42 to be expressed in
CC   mammalian cell culture and in transgenic mice.  The invention
CC   provides modified recombinant nucleic acid sequences and methods for
CC   increasing the mRNA levels and protein expression of proteins that
CC   are difficult to express in cell culture systems, mammalian cell
CC   culture systems or in transgenic animals.  The preferred difficult
CC   protein candidates for expression are those derived from lower
CC   organisms such as parasites, bacteria and viruses that have DNA
CC   coding sequences of high AT content or which have mRNA instability
CC   motifs or rare codons relative to the recombinant expression system
CC   to be used.  The invention allows expression of MSP-1 protein in
CC   the milk of transgenic animals, and also provides a DNA vaccine
CC   comprising a vector containing the altered MSP-1-42 sequence.
XX
SQ   Sequence   355 AA;


  Query Match             57.8%;  Score 1191.5;  DB 20;  Length 355;
  Best Local Similarity   60.2%;  Pred. No. 2.4e-76;
  Matches  222;  Conservative   54;  Mismatches   72;  Indels   21;  Gaps    2;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db       7 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 66

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db      67 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 126

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                 |:| :  |:|||| ||||||  :
Db     127 sekyksdldsikkyi------------------ndkqgenekylpflnnietlyktvn 166

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
           :||| ::|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db     167 dkidlfvihleakvlnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 226

Qy     246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDE 304
             ::| | ||||:| :|   |::| |::|  | |||||||||||||:||||||||||
Db     227 nhnnlltkflstgmvfenlaktvlsnlldgnlqgmlnisqhqcvkkqcpqnsgcfrhlde 286

Qy     305 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     287 reeckcllnykqegdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsy 346

Qy     365 PLFDGIFCS 373
           |||||||||
Db     347 plfdgifcs 355


RESULT   4
AAY09373
ID   AAY09373 standard; Protein; 361 AA.
XX
AC   AAY09373;
XX
DT   31-AUG-1999  (first entry)
XX
DE   Merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal.
XX
OS   Plasmodium falciparum.
XX
PN   WO9920774-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22226.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-288313/24.
DR   P-PSDB; AAX56009.
XX
PT   Modified malarial  protein for use in anti-malarial vaccines
XX
PS   Example; Fig 2; 35pp; English.
XX
CC   This present sequence comprises a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The C-terminal end of the sequence is
CC   modified to include a 6xHis tag.  A nucleic acid (see AAX56008)
CC   encoding MSP-1-42 has been modified according to a method
CC   of the invention in order to improve expression in mammalian cells
CC   and in transgenic animals by reducing the AT content and removing
CC   mRNA instability motifs.  The invention allows expression of
CC   MSP-1-42 in the milk of transgenic animals, and also provides a DNA
CC   vaccine comprising a vector containing the altered MSP-1-42 nucleic
CC   acid.
XX
SQ   Sequence   361 AA;


  Query Match             57.8%;  Score 1191.5;  DB 20;  Length 361;
  Best Local Similarity   60.2%;  Pred. No. 2.5e-76;
  Matches  222;  Conservative   54;  Mismatches   72;  Indels   21;  Gaps    2;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db       7 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 66
```

```
Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db      67 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 126

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db     127 sekyksdldsikkyi-------------------ndkqgenekylpflnnietlyktvn 166

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
           :||| ::|:|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db     167 dkidlfvihleakvlnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 226

Qy     246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDE 304
             ::| | ||||:| :|   |::| |::|  | ||||||||||||||:||||||||||
Db     227 nhnnlltkflstgmvfenlaktvlsnlldgnlqgmlnisqhqcvkkqcpqnsgcfrhlde 286

Qy     305 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     287 reeckcllnykqegdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsy 346

Qy     365 PLFDGIFCS 373
           |||||||||
Db     347 plfdgifcs 355
```

```
RESULT   5
AAY05833
ID   AAY05833 standard; Protein; 361 AA.
XX
AC   AAY05833;
XX
DT   02-AUG-1999  (first entry)
XX
DE   Merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal.
XX
OS   Plasmodium falciparum.
XX
PN   WO9920766-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22225.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-302742/25.
DR   N-PSDB; AAX25587.
XX
PT   New modified recombinant nucleic acid sequences useful for producing
PT   malarial DNA vaccine
XX
PS   Disclosure; Fig 2; 43pp; English.
XX
CC   This present sequence comprises a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The C-terminal end of the sequence is
CC   modified to include a 6xHis tag.  Nucleic acids (see AAX25586 and
CC   AAX25593) encoding MSP-1-42 have been modified according to a method
CC   of the invention in order to improve expression in mammalian cells
CC   and in transgenic animals.  The invention provides modified
CC   recombinant nucleic acid sequences and methods for increasing the
CC   mRNA levels and protein expression of proteins that are difficult
CC   to express in cell culture systems, especially mammalian cell
CC   culture systems or in transgenic animals.  The preferred difficult
CC   protein candidates for expression are those derived from lower
CC   organisms such as parasites, bacteria and viruses that have DNA
CC   coding sequences of high AT content or which have mRNA instability
CC   motifs or rare codons relative to the recombinant expression system
CC   to be used.  The invention allows expression of MSP-1 in the milk
CC   of transgenic animals, and also provides a DNA vaccine comprising a
CC   vector containing the altered MSP-1-42 nucleic acid.
XX
SQ   Sequence   361 AA;
```

```
  Query Match           57.8%;  Score 1191.5;  DB 20;  Length 361;
  Best Local Similarity   60.2%;  Pred. No. 2.5e-76;
  Matches  222;  Conservative   54;  Mismatches   72;  Indels   21;  Gaps    2;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db       7 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 66

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db      67 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 126

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db     127 sekyksdldsikkyi-------------------ndkqgenekylpflnnietlyktvn 166

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
           :||| ::|:|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db     167 dkidlfvihleakvlnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 226

Qy     246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDE 304
             ::| | ||||:| :|   |::| |::|  | ||||||||||||||:||||||||||
Db     227 nhnnlltkflstgmvfenlaktvlsnlldgnlqgmlnisqhqcvkkqcpqnsgcfrhlde 286

Qy     305 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     287 reeckcllnykqegdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsy 346

Qy     365 PLFDGIFCS 373
           |||||||||
Db     347 plfdgifcs 355
```

```
RESULT   6
AAY09374
ID   AAY09374 standard; Protein; 376 AA.
XX
AC   AAY09374;
XX
DT   31-AUG-1999  (first entry)
XX
DE   Modified merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal; mutant.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..15
FT                   /note= "beta-casein signal peptide"
FT   Peptide         371..376
FT                   /note= "6xHis tag"
FT   Misc-difference 197
FT                   /note= "Asn in native MSP-1-42 (N181Q mutation)"
FT   Misc-difference 278
FT                   /note= "Asn in native MSP-1-42 (N262Q mutation)"
XX
PN   WO9920774-A2.
```

```
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22226.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-288313/24.
DR   P-PSDB; AAX56008.
XX
PT   Modified malarial  protein for use in anti-malarial vaccines
XX
PS   Example; Fig 11; 35pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The sequence has been modified to include
CC   an N-terminal beta-casein signal peptide and an N-terminal 6xHis
CC   tag.  In addition, N181Q and N262Q mutations have been introduced to
CC   eliminate 2 N-glycosylation sites.  These modifications allow the
CC   MSP-1-42 protein to be expressed in the milk of transgenic mice.
XX
SQ   Sequence   376 AA;


  Query Match             57.4%;  Score 1184.5;  DB 20;  Length 376;
  Best Local Similarity   59.9%;  Pred. No. 8.1e-76;
  Matches  221;  Conservative   53;  Mismatches   74;  Indels   21;  Gaps    2;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db      22 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 81

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db      82 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 141

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db     142 sekyksdldsikkyi-------------------ndkqgenekylpflnnietlyktvn 181

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
           :||| ::|:|:||:     ||    ||| :|: || | ||:  || :|:|  |  |  |
Db     182 dkidlfvihleakvlqytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 241

Qy     246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDE 304
             ::| | ||||:| :|   |::| |::|  | || ||||||||||||:||||||||||
Db     242 nhnnlltkflstgmvfenlaktvlsnlldgnlqgmlqisqhqcvkkqcpqnsgcfrhlde 301

Qy     305 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 reeckcllnykqegdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsy 361

Qy     365 PLFDGIFCS 373
           |||||||||
Db     362 plfdgifcs 370


RESULT   7
AAY05834
ID   AAY05834 standard; Protein; 376 AA.
XX
AC   AAY05834;
XX
DT   02-AUG-1999  (first entry)
XX
DE   Modified merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal; mutant.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..15
FT                   /note= "beta-casein signal peptide"
FT   Peptide         371..376
FT                   /note= "6xHis tag"
FT   Misc-difference 197
FT                   /note= "Asn in native MSP-1-42 (N181Q mutation)"
FT   Misc-difference 278
FT                   /note= "Asn in native MSP-1-42 (N262Q mutation)"
XX
PN   WO9920766-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22225.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-302742/25.
DR   N-PSDB; AAX25593.
XX
PT   New modified recombinant nucleic acid sequences useful for producing
PT   malarial DNA vaccine
XX
PS   Example; Fig 11; 43pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The sequence has been modified to include
CC   an N-terminal beta-casein signal peptide and an N-terminal 6xHis
CC   tag.  In addition, N181Q and N262Q mutations have been introduced to
CC   eliminate 2 N-glycosylation sites.  These modifications allow the
CC   MSP-1-42 protein to be expressed in the milk of transgenic mice.
CC   The invention generally relates to modified recombinant nucleic
CC   acid sequences and methods for increasing the mRNA levels and
CC   protein expression of proteins that are difficult to express in
CC   cell culture systems, mammalian cell culture systems or in
CC   transgenic animals.  The preferred difficult protein candidates for
CC   expression are those derived from lower organisms such as parasites,
CC   bacteria and viruses that have DNA coding sequences of high AT
CC   content or which have mRNA instability motifs or rare codons
CC   relative to the recombinant expression system to be used.
XX
SQ   Sequence   376 AA;


  Query Match             57.4%;  Score 1184.5;  DB 20;  Length 376;
  Best Local Similarity   59.9%;  Pred. No. 8.1e-76;
  Matches  221;  Conservative   53;  Mismatches   74;  Indels   21;  Gaps    2;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db      22 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 81

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db      82 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 141

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
```

```
           ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db    142 sekyksdldsikkyi-------------------ndkqgenekylpflnnietlyktvn 181

Qy    186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
          :||| ::|:|:||:     ||    ||| :|: || | ||:  || :|:|  |  |  |
Db    182 dkidlfvihleakvlqytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 241

Qy    246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDE 304
            ::| | ||||:| :|   |::| |::|  | || ||||||||||||:|||||||||
Db    242 nhnnlltkflstgmvfenlaktvlsnlldgnlqgmlqisqhqcvkkqcpqnsgcfrhlde 301

Qy    305 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 364
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 reeckcllnykqegdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsy 361

Qy    365 PLFDGIFCS 373
          |||||||||
Db    362 plfdgifcs 370


RESULT  8
AAP50777
ID  AAP50777 standard; Protein; 1654 AA.
XX
AC  AAP50777;
XX
DT  30-SEP-1991  (first entry)
XX
DE  Sequence of the P195 protein of Plasmodium falciparum.
XX
KW  Malaria vaccine; epitope; antigen; immunogen.
XX
OS  Plasmodium falciparum.
XX
PN  EP154454-A.
XX
PD  11-SEP-1985.
XX
PF  21-FEB-1985;  85EP-0301173.
XX
PR  26-SEP-1984;  84GB-0024340.
PR  22-FEB-1984;  84GB-0004692.
PR  21-FEB-1985;  85GB-0004429.
XX
PA  (WELL ) WELLCOME FOUNDATION LTD.
XX
PI  Holder A,  Sandhu J,  Odink K,  Lockyer M,  Riveros-Moreno V;
XX
DR  WPI; 1985-224845/37.
DR  N-PSDB; AAN50530.
XX
PT  Cloned DNA sequence encoding plasmodium falciparum protein -
PT  useful for expressing the protein for use in vaccines against
PT  malaria
XX
PS  Claim 6; Fig 1; 51pp; English.
XX
CC  The sequence encoding the P195  protein of Plasmodium falciparum
CC  (AAN50530) and a peptide comprising at least one of its epitopes
CC  (see AAP50777) are claimed. Also claimed is a vaccine for inducing
CC  immunity to malaria comprising the novel peptide or P195 or a
CC  peptide comprising at least one epitope when derived from the new
CC  DNA sequence, together with a carrier.
XX
SQ  Sequence  1654 AA;


  Query Match           54.5%;  Score 1124;  DB 6;  Length 1654;
  Best Local Similarity  56.3%;  Pred. No. 1e-70;
  Matches  220;  Conservative   60;  Mismatches   89;  Indels   22;  Gaps    3;

Qy      6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
          :||||  ||||:|:||||||||||||||:| |: |||:|: ||||||  ||: | :||

Db   1284 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 1343

Qy     66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
          ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db   1344 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 1403

Qy    126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
            ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db   1404 sekyksdldsikkyi-------------------ndkqgenekylpflnnietlyktvn 1443

Qy    186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
          :||| ::|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db   1444 dkidlfvihleakvlnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 1503

Qy    246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLD 303
            ::| | ||||:| :|  :::| |:: |  : : : :     |    ::||||||||
Db   1504 nhnnlltkflstgmvfenllksvlsnlldwklaryvkhfttpmrkktmiqqssgcfrhld 1563

Qy    304 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDS 363
          |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db   1564 ereeckcllnykqegskcvensnptcnennggcdadakcteedsgsngkkitcectkpdc 1623

Qy    364 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||   |||||||||||||||||||||||||
Db   1624 yplsmvifcsssnflgisfllilmlilysfi 1654


RESULT  9
AAB37610
ID  AAB37610 standard; Protein; 262 AA.
XX
AC  AAB37610;
XX
DT  27-FEB-2001  (first entry)
XX
DE  Merozoite surface protein-133.
XX
KW  Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS  Plasmodium falciparum.
XX
PN  WO200063245-A2.
XX
PD  26-OCT-2000.
XX
PF  20-APR-2000; 2000WO-GB01558.
XX
PR  20-APR-1999;  99GB-0009072.
PR  13-MAY-1999;  99US-0311817.
PR  25-MAY-1999;  99CA-2271451.
XX
PA  (MEDI-) MEDICAL RES COUNCIL.
XX
PI  Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR  WPI; 2001-015762/02.
DR  N-PSDB; AAC68978.
XX
PT  Novel variants of the C-terminal fragment of Plasmodium merozoite
PT  surface protein-1, useful as vaccines for treating or preventing
PT  malaria -
XX
PS  Example 5; Fig 15; 126pp; English.
XX
CC  The present invention relates to non-natural variants of a C-terminal
CC  fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC  non-natural variants have reduced affinity for at least 1 antibody
CC  capable of blocking a second antibody that inhibits the proteolytic
CC  cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC  one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC  MSP-1_4_2, compared to natural MSP-1_1_9. The non-natural variants of the
CC  present invention are useful for immunising a mammal against malaria, and
CC  can be used to treat malaria. The present sequence is MSP-133 protein.
```

```
XX
SQ   Sequence   262 AA;


  Query Match             31.6%;  Score 651.5;  DB 22;  Length 262;
  Best Local Similarity   47.1%;  Pred. No. 1.9e-38;
  Matches  130;  Conservative   52;  Mismatches   73;  Indels   21;  Gaps    2;

Qy      6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
          :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db      7 idnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkrenfknvl 66

Qy     66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
          ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db     67 esdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvlgyykil 126

Qy    126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
            ||| ||:|||| |                    |:| :  |:|||| |||||  :
Db    127 sekyksdldsikkyi-------------------ndkqgenekylpflnnietlyktvn 166

Qy    186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
          :||| ::|:|:||:     ||    ||| :|: || | ||:  || :|:|  |  |  |
Db    167 dkidlfvihleakvlqytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdy 226

Qy    246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDML 280
            ::| | ||||:| :|   |::| |::|  | ||
Db    227 nhnnlltkflstgmvfenlaktvlsnlldgnlqgml 262


RESULT  10
AAW36103
ID   AAW36103 standard; Protein; 116 AA.
XX
AC   AAW36103;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)A protein sequence.
XX
KW   Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          1..95
FT                   /note= "amino acids derived from P. falciparum MSP1 p19
FT                         fragment"
FT   Region          96..116
FT                   /note= "glycosylphosphatidylinositol anchoring sequence"
XX
PN   WO9730158-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00290.
XX
PR   14-FEB-1996;   96FR-0001822.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425033/39.
DR   N-PSDB; AAT94550.
XX
PT   Recombinant protein containing the merozoite surface protein-1 p19
PT   fragment - useful in anti-malarial vaccines, diagnosis and protein
PT   purification
XX
PS   Disclosure; Fig 1B; 85pp; French.
XX
CC   This is the amino acid sequence of a recombinant protein comprising
CC   amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
CC   protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
CC   glycosylphosphatidylinositol membrane anchoring sequence.  p19 is the
CC   C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   116 AA;


  Query Match             31.0%;  Score 640;  DB 18;  Length 116;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-38;
  Matches  114;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    281 NISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3 nisqhqcvkkqcpensgcfrhldereeckcllnykqegdkcvenpnptcnennggcdada 62

Qy    341 KCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63 kcteedsgsngkkitcectkpdsyplfdgifcsssnflgisfllilmlilysfi 116


RESULT  11
AAW22592
ID   AAW22592 standard; Protein; 116 AA.
XX
AC   AAW22592;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)A protein sequence.
XX
KW   Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          1..95
FT                   /note= "amino acids derived from P. falciparum MSP1 p19
FT                         fragment"
FT   Region          96..116
FT                   /note= "glycosylphosphatidylinositol anchoring sequence"
XX
PN   WO9730159-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00291.
XX
PR   14-FEB-1996;   96FR-0001821.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425034/39.
DR   P-PSDB; AAW22592.
XX
PT   Recombinant protein containing Plasmodium merozoite surface
PT   protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT   antibodies for diagnosis and protein purification
XX
PS   Disclosure; Fig 1B; 85pp; French.
XX
```

```
CC   This is the amino acid sequence of a recombinant protein comprising
CC   amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
CC   protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
CC   glycosylphosphatidylinositol membrane anchoring sequence. p19 is the
CC   C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   116 AA;


  Query Match             31.0%;  Score 640;  DB 18;  Length 116;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-38;
  Matches  114;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     281 NISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3 nisqhqcvkkqcpensgcfrhldereeckcllnykqegdkcvenpnptcnennggcdada 62

Qy     341 KCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63 kcteedsgsngkkitcectkpdsyplfdgifcsssnflgisfllilmlilysfi 116


RESULT  12
AAB37609
ID   AAB37609 standard; Protein; 108 AA.
XX
AC   AAB37609;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Merozoite surface protein-119.
XX
KW   Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS   Plasmodium falciparum.
XX
PN   WO200063245-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-GB01558.
XX
PR   20-APR-1999;   99GB-0009072.
PR   13-MAY-1999;   99US-0311817.
PR   25-MAY-1999;   99CA-2271451.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR   WPI; 2001-015762/02.
DR   N-PSDB; AAC68977.
XX
PT   Novel variants of the C-terminal fragment of Plasmodium merozoite
PT   surface protein-1, useful as vaccines for treating or preventing
PT   malaria -
XX
PS   Example 5; Fig 15; 126pp; English.
XX
CC   The present invention relates to non-natural variants of a C-terminal
CC   fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC   non-natural variants have reduced affinity for at least 1 antibody
CC   capable of blocking a second antibody that inhibits the proteolytic
CC   cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC   one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC   MSP-1_4_2, compared to natural MSP-1_1_9. The non-natural variants of the
CC   present invention are useful for immunising a mammal against malaria, and
CC   can be used to treat malaria. The present sequence is MSP-119 protein.
XX
SQ   Sequence   108 AA;


  Query Match             26.9%;  Score 554;  DB 22;  Length 108;
  Best Local Similarity   92.4%;  Pred. No. 4.5e-32;
  Matches   97;  Conservative    3;  Mismatches    1;  Indels    4;  Gaps    1;

Qy     272 IEGKFQDMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNE 331
           |||:     ||:||||||||||:||||||||||||||||||||||||||||||||||||
Db       8 iegrh----niaqhqcvkkqcpqnsgcfrhldereeckcllnykqegdkcvenpnptcne 63

Qy     332 NNGGCDADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSN 376
           |||||||||||||||||||||||||||||||||||||||||||||
Db      64 nnggcdadakcteedsgsngkkitcectkpdsyplfdgifcsssn 108


RESULT  13
AAB37608
ID   AAB37608 standard; protein; 96 AA.
XX
AC   AAB37608;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Merozoite surface protein-1.
XX
KW   Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS   Plasmodium falciparum.
XX
PN   WO200063245-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-GB01558.
XX
PR   20-APR-1999;   99GB-0009072.
PR   13-MAY-1999;   99US-0311817.
PR   25-MAY-1999;   99CA-2271451.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR   WPI; 2001-015762/02.
XX
PT   Novel variants of the C-terminal fragment of Plasmodium merozoite
PT   surface protein-1, useful as vaccines for treating or preventing
PT   malaria -
XX
PS   Example 2; Page 48; 126pp; English.
XX
CC   The present invention relates to non-natural variants of a C-terminal
CC   fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC   non-natural variants have reduced affinity for at least 1 antibody
CC   capable of blocking a second antibody that inhibits the proteolytic
CC   cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC   one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC   MSP-1_4_2, compared to natural MSP-1_1_9. The present sequence is the
CC   wild-type MSP-1 protein. This sequence was used to generate the variants
CC   of the present invention. The non-natural variants of the present
CC   invention are useful for immunising a mammal against malaria, and can be
CC   used to treat malaria.
XX
SQ   Sequence   96 AA;


  Query Match             26.8%;  Score 553;  DB 22;  Length 96;
  Best Local Similarity   99.0%;  Pred. No. 4.5e-32;
  Matches   95;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     281 NISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 340
           |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db       1 nisqhqcvkkqcpqnsgcfrhldereeckcllnykqegdkcvenpnptcnennggcdada 60
```

```
Qy    341 KCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSN 376
          ||||||||||||||||||||||||||||||||||||
Db     61 kcteedsgsngkkitcectkpdsyplfdgifcsssn 96


RESULT  14
AAW22593
ID   AAW22593 standard; Protein; 127 AA.
XX
AC   AAW22593;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)S protein sequence.
XX
KW   Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Chimeric - Plasmodium vivax.
OS   Chimeric - Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /note= "signal peptide"
FT   Protein         20..127
FT                   /note= "mature protein"
FT   Region          1..32
FT                   /note= "derived from P. vivax MSP1"
FT   Region          33..34
FT                   /note= "encoded by restriction enzyme sequence used to
FT                          create the chimeric sequence"
FT   Region          35..127
FT                   /note= "derived from P. falciparum C-terminal p19
FT                          fragment of MSP1"
XX
PN   WO9730159-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00291.
XX
PR   14-FEB-1996;   96FR-0001821.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425034/39.
DR   P-PSDB; AAW22592.
XX
PT   Recombinant protein containing Plasmodium merozoite surface
PT   protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT   antibodies for diagnosis and protein purification
XX
PS   Disclosure; Fig 1C; 85pp; French.
XX
CC   This is the amino acid sequence of a chimeric protein comprising amino
CC   acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
CC   linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC   C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC   p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   127 AA;


  Query Match           26.5%;  Score 546;  DB 18;  Length 127;
  Best Local Similarity 97.9%;  Pred. No. 2e-31;
  Matches   94;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    278 DMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCD 337
          |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     32 defnisqhqcvkkqcpensgcfrhldereeckcllnykqegdkcvenpnptcnennggcd 91

Qy    338 ADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCS 373
          |||||||||||||||||||||||||||||||||||||
Db     92 adakcteedsgsngkkitcectkpdsyplfdgifcs 127


RESULT  15
AAW36102
ID   AAW36102 standard; Protein; 127 AA.
XX
AC   AAW36102;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)S protein sequence.
XX
KW   Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Chimeric - Plasmodium vivax.
OS   Chimeric - Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /note= "signal peptide"
FT   Protein         20..127
FT                   /note= "mature protein"
FT   Region          1..32
FT                   /note= "derived from P. vivax MSP1"
FT   Region          33..34
FT                   /note= "encoded by restriction enzyme sequence used to
FT                          create the chimeric sequence"
FT   Region          35..127
FT                   /note= "derived from P. falciparum C-terminal p19
FT                          fragment of MSP1"
XX
PN   WO9730158-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00290.
XX
PR   14-FEB-1996;   96FR-0001822.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425033/39.
DR   N-PSDB; AAT94549.
XX
PT   Recombinant protein containing the merozoite surface protein-1 p19
PT   fragment - useful in anti-malarial vaccines, diagnosis and protein
PT   purification
XX
PS   Disclosure; Fig 1C; 85pp; French.
XX
CC   This is the amino acid sequence of a chimeric protein comprising amino
CC   acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
CC   linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC   C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC   p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
```

SQ Sequence 127 AA;

Query Match 26.5%; Score 546; DB 18; Length 127;
Best Local Similarity 97.9%; Pred. No. 2e-31;
Matches 94; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy     278 DMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCD 337
           |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32 defnisqhqcvkkqcpensgcfrhldereeckcllnykqegdkcvenpnptcnennggcd 91

Qy     338 ADAKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCS 373
           ||||||||||||||||||||||||||||||||||||
Db      92 adakcteedsgsngkkitcectkpdsyplfdgifcs 127
```

Search completed: August 8, 2001, 12:31:53
Job time: 161 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:12 ; Search time 43.63 Seconds
(without alignments)
687.893 Million cell updates/sec

Title: US-09-500-376-2
Perfect score: 2062
Sequence: 1 AISVTMDNILSGFENEYDVI.....SNFLGISFLLILMLILYSFI 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2062	100.0	1726	1	SAZQGM	major merozoite su
2	2043.5	99.1	651	2	S47282	merozoite surface
3	2043	99.1	1726	2	A45948	major merozoite su
4	2027	98.3	1701	2	A54498	major merozoite su
5	2006	97.3	1701	2	A26868	major merozoite su
6	1289.5	62.5	400	2	A45545	major merozoite su
7	1289.5	62.5	1639	2	S05603	major merozoite su
8	1166	56.5	1631	1	SAZQK1	major merozoite su
9	1124	54.5	1640	2	A24594	probable major sur
10	843	40.9	1726	2	A39401	merozoite surface
11	832	40.3	1751	2	A45604	major blood-stage
12	613.5	29.8	680	2	A28121	major merozoite su
13	613.5	29.8	1772	2	A45532	major merozoite su
14	583	28.3	1785	2	A45546	major merozoite su
15	180.5	8.8	980	2	E71606	hypothetical prote
16	172	8.3	2166	2	G70163	hypothetical prote
17	162.5	7.9	1939	2	T18372	repeat organellar
18	159.5	7.7	1169	2	A64505	P115 homolog - Met
19	156	7.6	2401	2	T28676	rhoptry protein -
20	155.5	7.5	1619	2	T18499	hypothetical prote
21	153	7.4	1130	2	T34081	hypothetical prote
22	152.5	7.4	2269	2	T28677	rhoptry protein -
23	150	7.3	1156	2	B70356	chromosome assembl
24	145	7.0	3394	2	T18501	hypothetical prote
25	144.5	7.0	1086	2	S16752	major merozoite su
26	143	6.9	652	2	B59102	hypothetical prote
27	142	6.9	1250	2	E81339	probable restricti
28	141	6.8	442	2	T18507	hypothetical prote
29	141	6.8	1005	2	A64465	hypothetical prote
30	141	6.8	1127	2	T28317	ORF MSV156 hypothe
31	141	6.8	1187	2	T18355	hypothetical prote
32	140	6.8	624	2	PC6003	surface membrane p
33	139.5	6.8	671	2	H64502	hypothetical prote
34	139.5	6.8	1225	2	A49464	chromosome segrega
35	138.5	6.7	1979	2	C71622	hypothetical prote
36	138	6.7	1079	2	T18356	membrane protein p
37	138	6.7	1313	2	F96673	hypothetical prote
38	137.5	6.7	819	2	E70105	P115 protein homol
39	137	6.6	2116	2	A26655	myosin heavy chain
40	136.5	6.6	2829	2	A42771	reticulocyte-bindi
41	136	6.6	1679	2	S48385	hypothetical prote
42	136	6.6	2573	2	D71614	hypothetical prote
43	135	6.5	821	2	S67087	hypothetical prote
44	135	6.5	880	2	F75103	conserved hypothet
45	135	6.5	1228	2	A57384	multimerin, endoth

ALIGNMENTS

RESULT 1
SAZQGM
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N;Alternate names: 195K glycoprotein
C;Species: Plasmodium falciparum
C;Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
C;Accession: A23386; S06361
R;Weber, J.L.; Leininger, W.M.; Lyon, J.A.
Nucleic Acids Res. 14, 3311-3323, 1986
A;Title: Variation in the gene encoding a major merozoite surface antigen of the huma
A;Reference number: A23386; MUID:86205236
A;Accession: A23386
A;Molecule type: DNA
A;Residues: 1-1104 <WEB1>
A;Cross-references: EMBL:X03831
R;Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.
Nucleic Acids Res. 16, 1206, 1988
A;Title: Merozoite surface protein sequence from the Camp strain of the human malaria
A;Reference number: S06361; MUID:88143999
A;Accession: S06361
A;Molecule type: DNA
A;Residues: 1104-1726 <WEB2>
A;Cross-references: EMBL:X03831
C;Comment: The merozoite stages of different strains have strain-specific surface ant
C;Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1726/Product: major merozoite surface antigen #status predicted <MAT>
F;67-87,91-96,100-105,109-120/Region: 3-residue repeats (S-G-T)
F;757-765/Region: 3-residue repeats (T-E-E)
F;133,272,501,567,638,827,839,924,944,990,1016,1114,1221,1613,1658/Binding site: carb

Query Match 100.0%; Score 2062; DB 1; Length 1726;
Best Local Similarity 100.0%; Pred. No. 2.1e-100;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1333 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 1392

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1393 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1452

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1453 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1512

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1513 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 1572

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1573 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1632

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1633 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1692

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db   1693 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   2
S47282
merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71) (f
C;Species: Plasmodium falciparum
A;Variety: strain RO-71
C;Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 09-Jun-2000
C;Accession: S47282
R;Tolle, R.; Bujard, H.; Cooper, J.A.
submitted to the EMBL Data Library, July 1994
A;Description: Plasmodium falciparum: recombination within the C-terminal region of merc
A;Reference number: S47282
A;Accession: S47282
A;Molecule type: DNA
A;Residues: 1-651 <TOL>
A;Cross-references: EMBL:Z35329; NID:g535257; PIDN:CAA84558.1; PID:g535258
A;Experimental source: strain RO-71
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen


  Query Match             99.1%;  Score 2043.5;  DB 2;  Length 651;
  Best Local Similarity   99.5%;  Pred. No. 6.6e-100;
  Matches  392;  Conservative    1;  Mismatches    0;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    259 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 318

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 378

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    379 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 437

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    438 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 497

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db    498 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPQNSGCFR 557

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    558 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 617

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db    618 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 651


RESULT   3
A45948
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
C;Species: Plasmodium falciparum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jun-2000
C;Accession: A45948
R;Chang, S.P.; Kramer, K.J.; Yamaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.
Exp. Parasitol. 67, 1-11, 1988
A;Title: Plasmodium falciparum: gene structure and hydropathy profile of the major me
A;Reference number: A45948; MUID:89005525
A;Accession: A45948
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1726 <CHA>
A;Cross-references: GB:M37213
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match             99.1%;  Score 2043;  DB 2;  Length 1726;
  Best Local Similarity   99.2%;  Pred. No. 2e-99;
  Matches  391;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1333 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 1392

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1393 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1452

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          |||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||||||
Db   1453 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTTPSPVKTDEQKKESKFLPFLTNIETL 1512

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1513 YNNLVNKNDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 1572

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1573 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1632

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1633 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1692

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db   1693 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   4
A54498
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C;Accession: A54498
R;Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V
Mol. Biochem. Parasitol. 27, 291-302, 1988
A;Title: Variation in the precursor to the major merozoite surface antigens of Plasmo
A;Reference number: A54498; MUID:88142999
A;Accession: A54498
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1701 <PET>
A;Cross-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match             98.3%;  Score 2027;  DB 2;  Length 1701;
  Best Local Similarity   98.7%;  Pred. No. 1.4e-98;
  Matches  389;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1308 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 1367
```

```
Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1368 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1427

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1428 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1487

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    1488 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1547

Qy     241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1548 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1607

Qy     301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
           |||||||||||||||||||||||||||||||||||| |||||||| |||||||||
Db    1608 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 1667

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           ||||||||||||||||||||||||||||||||||
Db    1668 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1701


RESULT   5
A26868
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
C;Species: Plasmodium falciparum
C;Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
C;Accession: A26868
R;Tanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.
J. Mol. Biol. 195, 273-287, 1987
A;Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium
A;Reference number: A26868; MUID:88011243
A;Accession: A26868
A;Molecule type: DNA
A;Residues: 1-1701 <TAN>
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1701/Product: major merozoite surface antigen #status predicted <MAT>


  Query Match             97.3%;  Score 2006;  DB 2;  Length 1701;
  Best Local Similarity   98.0%;  Pred. No. 1.7e-97;
  Matches  386;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1308 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 1367

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    1368 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 1427

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1428 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1487

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    1488 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1547

Qy     241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    1548 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1607

Qy     301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
           |||||||||||||||||| |||||||||||||||| ||||||||: |||||||||
Db    1608 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 1667

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           ||||||||||||||||||||||||||||||||||
Db    1668 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1701


RESULT   6
A45545
major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments
C;Species: Plasmodium falciparum
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
C;Accession: A45545
R;Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.
Mol. Biochem. Parasitol. 49, 29-33, 1991
A;Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protei
A;Reference number: A45545; MUID:92131048
A;Accession: A45545
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-400 <BLA>
A;Note: sequence extracted from NCBI backbone (NCBIN:77612, NCBIP:77621)
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat


  Query Match             62.5%;  Score 1289.5;  DB 2;  Length 400;
  Best Local Similarity   62.3%;  Pred. No. 1e-60;
  Matches  243;  Conservative   54;  Mismatches   72;  Indels   21;  Gaps    2;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db      31 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 90

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db      91 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 150

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db     151 SEKYKSDLDSIKKYI-------------------NDKQGENEKYLPFLNNIETLYKTVN 190

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
           :||| ::|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db     191 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 250

Qy     246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDE 304
             ::| | ||||:| :|   |::| |::|  | ||||||||||||||:||||||||||
Db     251 NHNNLLTKFLSTGMVFENLAKTVLSNLLDGNLQGMLNISQHQCVKKQCPQNSGCFRHLDE 310

Qy     305 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 364
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 370

Qy     365 PLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||
Db     371 PLFDGIFCSSSNFLGISFLLILMLILYSFI 400


RESULT   7
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N;Alternate names: gp195 surface antigen
C;Species: Plasmodium falciparum
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C;Accession: S05603; S04850
R;Myler, P.J.
submitted to the EMBL Data Library, April 1989
A;Reference number: S05603
A;Accession: S05603
A;Molecule type: mRNA
A;Residues: 1-1639 <MYL>
A;Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
R;Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989
```

A;Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plasm
A;Reference number: S04850; MUID:89345116
A;Accession: S04850
A;Molecule type: mRNA
A;Residues: 1504-1639 <MYL2>
A;Cross-references: EMBL:X15063
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 62.5%; Score 1289.5; DB 2; Length 1639;
Best Local Similarity 62.3%; Pred. No. 4.9e-60;
Matches 243; Conservative 54; Mismatches 72; Indels 21; Gaps 2;

Qy 6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
Db 1270 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 1329

Qy 66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
Db 1330 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 1389

Qy 126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
Db 1390 SEKYKSDLDSIKKYI----------NDKQGENEKYLPFLNNIETLYKTVN 1429

Qy 186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
Db 1430 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 1489

Qy 246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDE 304
Db 1490 NHNNLLTKFLSTGMVFENLAKTVLSNLLDGNLQGMLNISQHQCVKKQCPQNSGCFRHLDE 1549

Qy 305 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 364
Db 1550 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 1609

Qy 365 PLFDGIFCSSSNFLGISFLLILMLILYSFI 394
Db 1610 PLFDGIFCSSSNFLGISFLLILMLILYSFI 1639

RESULT 8
SAZQK1
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
C;Species: Plasmodium falciparum
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000
C;Accession: A25120
R;Mackay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.; E
EMBO J. 4, 3823-3829, 1985
A;Title: Polymorphism of the precursor for the major surface antigens of Plasmodium falc
A;Reference number: A91030; MUID:86136024
A;Accession: A25120
A;Molecule type: DNA
A;Residues: 1-1631 <MAC>
C;Comment: The merozoite stages of different strains have strain-specific surface antige
C;Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The me
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane prote
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1631/Product: major merozoite surface antigen #status predicted <MAT>
F;67-84/Region: 3-residue repeats (S-G-T/P)
F;1614-1631/Domain: membrane anchor #status predicted <MBN>
F;97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (Asn) (

Query Match 56.5%; Score 1166; DB 1; Length 1631;
Best Local Similarity 58.1%; Pred. No. 1.4e-53;
Matches 227; Conservative 57; Mismatches 85; Indels 22; Gaps 3;

Qy 6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
Db 1261 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 1320

Qy 66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
Db 1321 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 1380

Qy 126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
Db 1381 SEKYKSDLDSIKKYI----------NDKQGENEKYLPFLNNIETLYKTVN 1420

Qy 186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
Db 1421 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELIYLKTIQDKLADFKKNNNFVGIADLSTDY 1480

Qy 246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLD 303
Db 1481 NHNNLLTKFLSTGMVFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLD 1540

Qy 304 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDS 363
Db 1541 EREECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDS 1600

Qy 364 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
Db 1601 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 1631

RESULT 9
A24594
probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodi
C;Species: Plasmodium falciparum
C;Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jun-2000
C;Accession: A24594
R;Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholl
Nature 317, 270-273, 1985
A;Title: Primary structure of the precursor to the three major surface antigens of Pl
A;Reference number: A24594; MUID:86014355
A;Accession: A24594
A;Molecule type: DNA
A;Residues: 1-1640 <HOL>
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

Query Match 54.5%; Score 1124; DB 2; Length 1640;
Best Local Similarity 56.3%; Pred. No. 2.2e-51;
Matches 220; Conservative 60; Mismatches 89; Indels 22; Gaps 3;

Qy 6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
Db 1270 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 1329

Qy 66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
Db 1330 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 1389

Qy 126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
Db 1390 SEKYKSDLDSIKKYI----------NDKQGENEKYLPFLNNIETLYKTVN 1429

Qy 186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
Db 1430 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 1489

Qy 246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRHLD 303
Db 1490 NHNNLLTKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLD 1549

Qy 304 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDS 363
Db 1550 EREECKCLLNYKQEGSKCVENSNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDC 1609

```
Qy    364 YPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||   |||||||||||||||||||||||||
Db   1610 YPLSMVIFCSSSNFLGISFLLILMLILYSFI 1640
```

RESULT 10
A39401
merozoite surface antigen 1 precursor - Plasmodium vivax
C;Species: Plasmodium vivax
C;Date: 28-Feb-1992 #sequence_revision 28-Feb-1992 #text_change 24-Nov-1999
C;Accession: A39401
R;del Portillo, H.A.; Longacre, S.; Khouri, E.; David, P.H.
Proc. Natl. Acad. Sci. U.S.A. 88, 4030-4034, 1991
A;Title: Primary structure of the merozoite surface antigen 1 of Plasmodium vivax reveal
A;Reference number: A39401; MUID:91219506
A;Accession: A39401
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1726 <DEL>
A;Cross-references: GB:M60807
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

```
  Query Match             40.9%;  Score 843;  DB 2;  Length 1726;
  Best Local Similarity   43.7%;  Pred. No. 1.2e-36;
  Matches  169;  Conservative   81;  Mismatches  115;  Indels   22;  Gaps    7;

Qy     15 NEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLESDLMQFKH 74
          ::|||:|||||||:|:  :|||:| ::  || |: |:|:|||||| |||:|| |||  ||:
Db   1346 SDYDVVYLKPLAGMYKKIKKQLENHVNAFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKY 1405

Qy     75 ISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLAKYKDDL- 133
           | ||||:| :|||: |:|  || |||||  |:: |:  | :|::|| |:   ||  |
Db   1406 SPSGEYIIKDPYKLLDLEKKKKLLGSYKYIGASIDKDLGTANDGVNYYNKMGELYKTHLT 1465

Qy    134 ---ESIKKV---IKEEKEKFPSSPPTTPPSPAKTDEQKKES---KFLPFLTNIETLYNNL 184
             | :|||   || | :|       :  :  ||    |::   |:|||| :::  | :|
Db   1466 AVNEEVKKVEADIKAEDDKIKKIGSDSTKTTEKTQSMAKKAELEKYLPFLNSLQKEYESL 1525

Qy    185 VNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLIND 244
          |:|:: |  |||  ||:| :|| || : : || |   :|:|:: :|     :: ||  |:
Db   1526 VSKVNTYTDNLKKVINNCQLEKKEAEITVKKLQDYNKMDEKLEEYK-----KSEKK--NE 1578

Qy    245 DTKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLD 303
             :| ||: : |: :|    |:|:|:  : | :   |:| |:    |:|: |:|:||
Db   1579 VKSSGLLEKLMKSKLIKENESKEILSQLLNVQTQLLTMSSEHTCIDTNVPDNAACYRYLD 1638

Qy    304 EREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDS 363
             || :||| :|:|| |||   | || :|||||  :|:|   ||     || |:|||  |
Db   1639 GMEEWRCLLTFKEEGGKCVPGSNVTCKDNNGGCAPEAECKMTDS----NKIVCKCTKEGS 1694

Qy    364 YPLFDGIFCSSSNFLGISFLLILMLIL 390
           |||:|:|||||:|| :||||:::| |
Db   1695 EPLFEGVFCSSSSFLSLSFLLLMLLFL 1721

RESULT 11
A45604
major blood-stage surface antigen Pv200 - Plasmodium vivax
C;Species: Plasmodium vivax
C;Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
C;Accession: A45604
R;Gibson, H.L.; Tucker, J.E.; Kaslow, D.C.; Krettli, A.U.; Collins, W.E.; Kiefer, M.C.;
Mol. Biochem. Parasitol. 50, 325-333, 1992
A;Title: Structure and expression of the gene for Pv200, a major blood-stage surface ant
A;Reference number: A45604; MUID:92158013
A;Accession: A45604
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1751 <GIB>
A;Cross-references: GB:M75674; NID:g160608; PID:g457336
A;Note: sequence extracted from NCBI backbone (NCBIN:83591, NCBIP:83592)
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

```
  Query Match   40.3%; Score 832; DB 2; Length 1751;
  Best Local Similarity   41.5%; Pred. No. 4.4e-36;
  Matches 165; Conservative   84; Mismatches 105; Indels   44; Gaps   6;

Qy   15 NEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLESDLMQFKH 74
          ::|||:|||||||:|:::|||:| :: || |: |:|:|||||| |||:|| ||| ||:
Db   1371 SDYDVVYLKPLAGMYKTIKKQLENHVNAFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKY 1430

Qy   75 ISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLAKYKDDLE 134
           || ||||:| :|||: |:| |: ||||| |:: |: | :|::|| |: || |:
Db   1431 SSSGEYIIKDPYKLLDLEKKKKLIGSYKYIGASIDMDLATANDGVTYYNKMGELYKTHLD 1490

Qy   135 SIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKE----------------------SKFLPF 173
           :| ||: :: | ||: |: |:|||
Db   1491 GVKTEIKKVEDDI-----------KKQDEELKKLGNVNSQDSKKNEFIAKKAELEKYLPF 1539

Qy   174 LTNIETLYNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHN 233
          | ::: | :||:|:: | ||| ||:| :|| || : : || | :|:|:: :|
Db   1540 LNSLQKEYESLVSKVNTYTDNLKKVINNCQLEKKEAEITVKKLQDYNKMDEKLEEYK--- 1596

Qy   234 DFEAIKKLINDDTKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNISQHQCVKKQC 292
            :: || |:   :| ||: : |: :| |:|:|: : | : |:| |:
Db   1597 --KSEKK--NEVKSSGLLEKLMKSKLIKENESKEILSQLLNVQTQLLTMSSEHTCIDTNV 1652

Qy   293 PENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGK 352
          |:|: |:|:|| || :||| :|:|| ||| | || :||||| :|:| ||
Db   1653 PDNAACYRYLDGTEEWRCLLTFKEEGGKCVPASNVTCKDNNGGCAPEAECKMTDS----N 1708

Qy   353 KITCECTKPDSYPLFDGIFCSSSNFLGISFLLILMLIL 390
          || |:||| | |||:|:||||||:|| :||||:::| |
Db   1709 KIVCKCTKEGSEPLFEGVFCSSSSFLSLSFLLLMLLFL 1746

RESULT 12
A28121
major merozoite surface antigen - Plasmodium yoelii (fragment)
C;Species: Plasmodium yoelii
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 26-Aug-1999
C;Accession: A28121
R;Burns Jr., J.M.; Daly, T.M.; Vaidya, A.B.; Long, C.A.
Proc. Natl. Acad. Sci. U.S.A. 85, 602-606, 1988
A;Title: The 3' portion of the gene for a Plasmodium yoelii merozoite surface antigen
A;Reference number: A28121; MUID:88124889
A;Accession: A28121
A;Molecule type: DNA
A;Residues: 1-680 <BUR>
A;Cross-references: GB:J03612; NID:g160678; PID:g160679
A;Experimental source: strain 17XL
A;Note: the authors translated the codon GTA for residue 429 as Leu
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

```
  Query Match             29.8%;  Score 613.5;  DB 2;  Length 680;
  Best Local Similarity   33.5%;  Pred. No. 4.2e-25;
  Matches  131;  Conservative   90;  Mismatches  143;  Indels   27;  Gaps    7;

Qy      8 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLES 67
          :||| | ||   :| | |   |:|||| : :   |   :: : ||:: :||  ||:||
Db    309 DILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQKRNDFLEVLSH 368

Qy     68 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 127
           :|  || :|:|:|:| : ::||::::|:  : : ||  : :  ||:   :|| :: |::
Db    369 ELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFNKMVE 428

Qy    128 KYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKE--SKFLPFLTNIETLYNNLV 185
           |   | ::|: |   :                |:|::  ||   |::| | ::: ||  ::
Db    429 VYNTQLAAVKEQIATIE-------------AETNDTNKEEKKKYIPILEDLKGLYETVI 474
```

```
Qy    186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
          : ::|   |: ::::   || |  :    |     ||:|:| |  | :         |
Db    475 GQAEEYSEELQNRLDNYKNEKAEFEILTKNLEKYIQIDEKLDEFVEHAE-------NNKH 527

Qy    246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNIS-QHQCV-KKQCPENSGCFRHL 302
              |  |  :||| :     |::|::     |:| :  :| ||  :  |:|:|||| |
Db    528 IASIALNNLNKSGLVGEGESKKILAKMLNMDGMDLLGVDPKHVCVDTRDIPKNAGCFRDD 587

Qy    303 DEREECKCLLNYKQ-EGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKP 361
          :  || :||| ||: ||: |||| |||| |||||  | |   :|  | ||| | | :|
Db    588 NGTEEWRCLLGYKKGEGNTCVENNNPTCDINNGGCDPTASCQNAESTENSKKIICTCKEP 647

Qy    362 DSYPLFDGIFCSSSNFLGISFLLILMLILYS 392
              ::|:|||||:|:|:| |||: ||:::
Db    648 TPNAYYEGVFCSSSSFMGLSILLIITLIVFN 678


RESULT  13
A45532
major merozoite surface antigen precursor - Plasmodium yoelii
C;Species: Plasmodium yoelii
C;Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 26-Aug-1999
C;Accession: A45532; A45531
R;Lewis, A.P.
Mol. Biochem. Parasitol. 36, 271-282, 1989
A;Title: Cloning and analysis of the gene encoding the 230-kilodalton merozoite surface
A;Reference number: A45532; MUID:90014981
A;Accession: A45532
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1772 <LEW>
A;Cross-references: GB:J04668; NID:g160492; PID:g160493
R;Daly, T.M.; Burns Jr., J.M.; Long, C.A.
Mol. Biochem. Parasitol. 36, 283-285, 1989
A;Title: Precursor to the major merozoite surface antigen of Plasmodium yoelii: cloning
A;Reference number: A45531; MUID:90014982
A;Accession: A45531
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 454-1094 <DAL>
A;Cross-references: GB:J03975; NID:g160081; PID:g160082
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match             29.8%;  Score 613.5;  DB 2;  Length 1772;
  Best Local Similarity   33.5%;  Pred. No. 1.2e-24;
  Matches  131;  Conservative   90;  Mismatches  143;  Indels   27;  Gaps    7;

Qy      8 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLES 67
          :||| | ||   :| | |   |:|||| : :   |   :: : ||:: :|| ||:||
Db   1401 DILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQKRNDFLEVLSH 1460

Qy     68 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 127
          :|  || :|:|:|:| : ::||:::|:  : : ||  : :  ||:   :|| :: |::
Db   1461 ELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFNKMVE 1520

Qy    128 KYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKE--SKFLPFLTNIETLYNNLV 185
            |   | ::|: |   :              |:|::  ||   |::| | ::: ||  ::
Db   1521 LYNTQLAAVKEQIATIE-------------AETNDTNKEEKKKYIPILEDLKGLYETVI 1566

Qy    186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
          : ::|   |: ::::   || |  :    |     ||:|:| |  | :         |
Db   1567 GQAEEYSEELQNRLDNYKNEKAEFEILTKNLEKYIQIDEKLDEFVEHAE-------NNKH 1619

Qy    246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNIS-QHQCV-KKQCPENSGCFRHL 302
              |  |  :||| :     |::|::     |:| :  :| ||  :  |:|:|||| |
Db   1620 IASIALNNLNKSGLVGEGESKKILAKMLNMDGMDLLGVDPKHVCVDTRDIPKNAGCFRDD 1679

Qy    303 DEREECKCLLNYKQ-EGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKP 361
          :  || :||| ||: ||: |||| |||| |||||  | |   :|  | ||| | | :|
Db   1680 NGTEEWRCLLGYKKGEGNTCVENNNPTCDINNGGCDPTASCQNAESTENSKKIICTCKEP 1739

Qy    362 DSYPLFDGIFCSSSNFLGISFLLILMLILYS 392
              ::|:|||||:|:|:| |||: ||:::
Db   1740 TPNAYYEGVFCSSSSFMGLSILLIITLIVFN 1770


RESULT  14
A45546
major merozoite surface antigen precursor - Plasmodium chabaudi chabaudi
C;Species: Plasmodium chabaudi chabaudi
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
C;Accession: A45546
R;Deleersnijder, W.; Hendrix, D.; Bendahman, N.; Hanegreefs, J.; Brijs, L.; Hamers-Ca
Mol. Biochem. Parasitol. 43, 231-244, 1990
A;Title: Molecular cloning and sequence analysis of the gene encoding the major meroz
A;Reference number: A45546; MUID:91218805
A;Accession: A45546
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1785 <DEL>
A;Cross-references: GB:M34947; NID:g160597; PID:g160598
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match             28.3%;  Score 583;  DB 2;  Length 1785;
  Best Local Similarity   33.2%;  Pred. No. 4.8e-23;
  Matches  131;  Conservative   92;  Mismatches  137;  Indels   34;  Gaps   11;

Qy      8 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDI---LNSRLKKRKYFLDV 64
          :|| |::| : || | |   |:| || :.|    |::   ||   || :|:|:|  ||||
Db   1415 DILDAFKSENEYIYTKSLGNTYKSFKKHMLKE---FSMIKEDIMTGLNYKLEKRNDFLDV 1471

Qy     65 LESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEK 124
          |  :|  || |::|:::::: ::||::::|:  : : ||  : |  ||: | :|| :: |
Db   1472 LSYELALFKDINTNKFVVKNPYQLLDNDKKDKQMINLKYAIKGVTEDIETATDGIEFFNK 1531

Qy    125 VLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNL 184
          ::  ||  | :: : |         ::  | |     || :||  |: |   ::: ||  :
Db   1532 MIELYKPQLNAVNEQI--------AAIGTEP-----TDAEKK--KYAPIFEDLKGLYETI 1576

Qy    185 VNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLIND 244
          :|  :::   |: |: :  :||    : :  |     ||:|:: |       |: :|  |
Db   1577 LNGAEEFSELLQHKLENYKIEKAGFDILMANLETYIRIDEKLEDF-----VESAEK--NK 1629

Qy    245 DTKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNI-SQHQCVKKQCPENSGCFRHL 302
              |  |  :||| :     |::|::     |:| | | | |:    |:|:||||:
Db   1630 HIASIALNNLNKSGLVTEGESKKILAKMLNMDAMDLLGIGSNHVCISTSTPDNAGCFRYD 1689

Qy    303 DEREECKCLLNYKQ--EGDKCVENPNPTCNENNGGCDADAKCTEEDSGSN--GKKITCEC 358
          |  || :||| :|:  :|::|| :  | || ||||||  :| | | ::       ||| | |
Db   1690 DGTEEWRCLLGFKKDDDGNRCVADDAPVCNNNNGGCDKNADCREVENTDRDPSKKIVCTC 1749

Qy    359 TKPDSYPLFDGIFCSSSNFLGISFLLILMLILYS 392
           :|:    : |:|||| |:|:| |||: ||:::
Db   1750 KEPNPNAYYAGVFCSSSGFMGLSILLIITLIVFN 1783


RESULT  15
E71606
hypothetical protein PFB0765w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: E71606
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743
A;Accession: E71606
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-980 <GAR>
```

```
A;Cross-references: GB:AE001417; GB:AE001362; NID:g3845271; PIDN:AAC71942.1; PID:g384527
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0765w
C;Superfamily: hypothetical protein MJ1322


  Query Match             8.8%;  Score 180.5;  DB 2;  Length 980;
  Best Local Similarity   24.0%;  Pred. No. 0.028;
  Matches  103;  Conservative   75;  Mismatches  141;  Indels  111;  Gaps   23;

Qy      14 ENEYDVIYLKPLAGVYRSLKKQIEKNI-------------FTFNLNLNDIL---NSRLKK 57
           :| |   |||     |    |::|||              :  |  ||: :   || :|
Db     248 QNYY---YLKK---EYDLKNKELEKNIEHGKKLEHELSHCYEENQKLNEEIKRRNSFIKN 301

Qy      58 RKYFLDVL---ESDLMQFKHIS------------SNEYIIED--------------SFK 87
           :   :|:|   |::|:: | |:             :|| :::|                :
Db     302 KDRKIDLLTNIENELLKKKEINNIKLMEKQNVIKNNEQLLKDIKDENEKMNEHVNKLQNE 361

Qy      88 LLNSEQKNTLL-KSYKYIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKK-------- 138
           |:  | :|  : |  :: |:  |: ||  :: :   :| :   ||:| :|||
Db     362 LIKRELQNKCISKDIEFCKKEKEDKIKNLEDDLLEKKKCIENLKDELINIKKKMEDKMHM 421

Qy     139 -----VIKEEKEKFPSSPPTTPPSPAKTDEQ----KKESKFLPFLTNIETLYN-NLVNKI 188
                ::   : |:       |   :  : : :    ||:     ||   |   | ::| ||
Db     422 TNEMDLLSNKVEELNRINKTYEKNIVELNNELDVIKKKLNDEEFLKEEEKKKNIDMVYKI 481

Qy     189 DDYLINLKAKINDC-NVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDDTK 247
           :| | :| | |:  :::|:| :: : |  :|   ::|  : ||  | | |  :|    |
Db     482 KEYEIQIKEKENEIDSLKKNEQNLHVLKNEEL---NEKEIILKNKYDKE-INMIIEQYNK 537

Qy     248 -----KDMLGKLLSTGLVQNFPNTIISKLIEGKFQ--------DMLNISQHQCVKKQCPE 294
                ||||   : : : |   | |     | | :        || |:     :|    |
Db     538 KIQEEKDMLNNKIKS-MDQTHKNQIEEMQEENKKELKRLKNVCDM-NLQSQILIK----E 591

Qy     295 NSGCFRHLDER-EECKCLLNYK--------QEGDKCVENPNPTCNENNGGCDADAKCTEE 345
           |    :|: |: || | ||  |         || |: :|  |    :    |:   |
Db     592 NE---KHMQEKVEEYKNLLKQKDQELKNIIQEYDERIEIQNKEMEDIVNDCEEKLK---- 644

Qy     346 DSGSNGKKIT 355
           :  | ||:|
Db     645 QAKINNKKLT 654


Search completed: August  8, 2001, 12:30:49
Job time: 97 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:12 ; Search time 24.79 Seconds
(without alignments)
544.440 Million cell updates/sec

Title: US-09-500-376-2
Perfect score: 2062
Sequence: 1 AISVTMDNILSGFENEYDVI.....SNFLGISFLLILMLILYSFI 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2062	100.0	1726	1	MSP1_PLAFC	P04934 plasmodium
2	2059	99.9	1726	1	MSP1_PLAFP	P50495 plasmodium
3	2027	98.3	1701	1	MSP1_PLAFF	P13819 plasmodium
4	2027	98.3	1701	1	MSP1_PLAFM	P08569 plasmodium
5	2025.5	98.2	1682	1	MSP1_PLAF3	P19598 plasmodium
6	1289.5	62.5	1630	1	MSP1_PLAFK	P04932 plasmodium
7	1289.5	62.5	1639	1	MSP1_PLAFW	P04933 plasmodium
8	613.5	29.8	1772	1	MSP1_PLAYO	P13828 plasmodium
9	153	7.4	1130	1	YL17_CAEEL	Q11102 caenorhabdi
10	139.5	6.8	1225	1	SMC1_YEAST	P32908 saccharomyc
11	137	6.6	2116	1	MYS2_DICDI	P08799 dictyosteli
12	136.5	6.6	2869	1	RBP1_PLAVB	Q00798 plasmodium
13	136	6.6	1679	1	YIO9_YEAST	P40457 saccharomyc
14	135	6.5	1228	1	ECM_HUMAN	Q13201 homo sapien
15	130.5	6.3	1790	1	USO1_YEAST	P25386 saccharomyc
16	129	6.3	944	1	NUF1_YEAST	P32380 saccharomyc
17	129	6.3	2245	1	MYSJ_DICDI	P54697 dictyosteli
18	128	6.2	796	1	YFC3_YEAST	P43573 saccharomyc
19	128	6.2	1038	1	CIN8_YEAST	P27895 saccharomyc
20	128	6.2	1251	1	RBP2_PLAVB	Q00799 plasmodium
21	127	6.2	540	1	YKZ6_YEAST	P36112 saccharomyc
22	127	6.2	729	1	KAR3_YEAST	P17119 saccharomyc
23	127	6.2	1324	1	CUT3_SCHPO	P41004 schizosacch
24	126	6.1	1312	1	RA50_YEAST	P12753 saccharomyc
25	125.5	6.1	622	1	YAE7_YEAST	P39723 saccharomyc
26	125	6.1	471	1	SG3_MOUSE	P47867 mus musculu
27	124.5	6.0	1875	1	MLP1_YEAST	Q02455 saccharomyc
28	124	6.0	1928	1	MYS1_YEAST	P08964 saccharomyc
29	123.5	6.0	703	1	HS83_PHANI	P51819 pharbitis n
30	123.5	6.0	798	1	ITB1_HUMAN	P05556 homo sapien
31	122.5	5.9	1230	1	SMC3_YEAST	P47037 saccharomyc
32	122	5.9	399	1	RMS5_PENUR	P47907 penicillium
33	122	5.9	1170	1	SMC2_YEAST	P38989 saccharomyc
34	122	5.9	1805	1	HMW2_MYCGE	P47460 mycoplasma
35	121.5	5.9	506	1	CC37_YEAST	P06101 saccharomyc
36	121.5	5.9	984	1	NOF_DROME	P16320 drosophila
37	121	5.9	569	1	CYSP_PLAFA	P25805 plasmodium
38	121	5.9	803	1	ITB1_CHICK	P07228 gallus gall
39	121	5.9	1101	1	DIA2_HUMAN	O60879 homo sapien
40	120.5	5.8	1526	1	MYS2_SCHPO	Q9usi6 schizosacch
41	120.5	5.8	1853	1	MY5A_MOUSE	Q99104 mus musculu
42	120	5.8	743	1	ABRA_PLAFC	P22620 plasmodium
43	119.5	5.8	400	1	OMPA_THEMA	Q01969 thermotoga
44	119.5	5.8	600	1	ABRA_PLAFG	P23745 plasmodium
45	119.5	5.8	633	1	YHR4_YEAST	P38822 saccharomyc

ALIGNMENTS

```
RESULT   1
MSP1_PLAFC
ID   MSP1_PLAFC     STANDARD;      PRT;  1726 AA.
AC   P04934;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P195).
GN   MSP-1.
OS   Plasmodium falciparum (isolate Camp / Malaysia).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5835;
RN   [1]
RP   SEQUENCE OF 1-1103 FROM N.A.
RX   MEDLINE=86205236; PubMed=3517809;
RA   Weber J.L., Leininger W.M., Lyon J.A.;
RT   "Variation in the gene encoding a major merozoite surface antigen of
RT   the human malaria parasite Plasmodium falciparum.";
RL   Nucleic Acids Res. 14:3311-3323(1986).
RN   [2]
RP   SEQUENCE OF 1104-1726 FROM N.A.
RX   MEDLINE=88143999; PubMed=3278296;
RA   Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
RT   "Merozoite surface protein sequence from the Camp strain of the human
RT   malaria parasite Plasmodium falciparum.";
RL   Nucleic Acids Res. 16:1206-1206(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X03831; CAA27446.1; -.
DR   PIR; A23386; SAZQGM.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1726       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    133    133       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    567    567       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    638    638       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    924    924       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    944    944       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    990    990       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1016   1016       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1114   1114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1221   1221       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1613   1613       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1726 AA;  196197 MW;  DD8AD45FA352BCF3 CRC64;


  Query Match           100.0%;  Score 2062;  DB 1;  Length 1726;
  Best Local Similarity  100.0%;  Pred. No. 9e-99;
  Matches  394;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1333 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 1392

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1393 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1452

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1453 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1512

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1513 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 1572

Qy     241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1573 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1632

Qy     301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1633 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1692

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db    1693 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   2
MSP1_PLAFP
ID   MSP1_PLAFP     STANDARD;      PRT;  1726 AA.
AC   P50495;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (GP195).
GN   MSP-1.
OS   Plasmodium falciparum (isolate Palo Alto / Uganda).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=57270;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89005525; PubMed=3049134;
RA   Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
RA   Siddiqui W.A.;
RT   "Plasmodium falciparum: gene structure and hydropathy profile of the
RT   major merozoite surface antigen (gp195) of the Uganda-Palo Alto
RT   isolate.";
RL   Exp. Parasitol. 67:1-11(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M37213; AAA29611.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1726       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD   133    133       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   567    567       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   638    638       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   924    924       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   944    944       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   990    990       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1016   1016       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1114   1114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1221   1221       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1613   1613       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE  1726 AA;  196174 MW;  5B59CEEFA2F9A026 CRC64;


  Query Match           99.9%;  Score 2059;  DB 1;  Length 1726;
  Best Local Similarity  99.7%;  Pred. No. 1.3e-98;
  Matches  393;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1333 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 1392

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1393 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1452

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1453 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1512

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db    1513 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFDAIKK 1572

Qy     241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1573 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1632

Qy     301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1633 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1692

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db    1693 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   3
MSP1_PLAFF
ID   MSP1_PLAFF     STANDARD;      PRT;  1701 AA.
AC   P13819;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA).
GN   MSP-1.
OS   Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```

```
OX   NCBI_TaxID=5837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88142999; PubMed=2449612;
RA   Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
RA   Brown G.V., Anders R.F., Kemp D.J.;
RT   "Variation in the precursor to the major merozoite surface antigens
RT   of Plasmodium falciparum.";
RL   Mol. Biochem. Parasitol. 27:291-302(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M19143; AAA29653.1; -.
DR   PIR; A54498; A54498.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1701       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    536    536       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    607    607       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    802    802       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    899    899       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    965    965       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    991    991       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1089   1089       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1196   1196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1588   1588       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1701 AA;  193719 MW;  3920B75E73D38552 CRC64;
```

```
  Query Match           98.3%;  Score 2027;  DB 1;  Length 1701;
  Best Local Similarity  98.7%;  Pred. No. 5.5e-97;
  Matches  389;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1308 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 1367

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1368 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1427

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1428 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1487

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    1488 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1547

Qy     241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1548 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1607

Qy     301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
           |||||||||||||||||||||||||||||||||||| |||||||: |||||||||
Db    1608 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 1667

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db    1668 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1701
```

```
RESULT   4
MSP1_PLAFM
ID   MSP1_PLAFM     STANDARD;      PRT;  1701 AA.
AC   P08569;
DT   01-AUG-1988 (Rel. 08, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P190).
GN   MSP-1.
OS   Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=70153;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88011243; PubMed=3079521;
RA   Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT   "Allelic dimorphism in a surface antigen gene of the malaria parasite
RT   Plasmodium falciparum.";
RL   J. Mol. Biol. 195:273-287(1987).
RN   [2]
RP   REVISIONS TO 1403; 1569 AND 1629.
RA   Tanabe K.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-115 FROM N.A.
RX   MEDLINE=86136024; PubMed=3004972;
RA   Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA   Stunnenberg H., Bujard H.;
RT   "Polymorphism of the precursor for the major surface antigens of
RT   Plasmodium falciparum merozoites: studies at the genetic level.";
RL   EMBO J. 4:3823-3829(1985).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X05624; CAA29112.1; -.
DR   PIR; A26868; A26868.
DR   PIR; B25120; B25120.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1701       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    536    536       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    607    607       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    802    802       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    899    899       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    965    965       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    991    991       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1089   1089       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD   1196   1196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1588   1588       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1701 AA;  193768 MW;  3FC2EC59AF96EA98 CRC64;


  Query Match            98.3%;  Score 2027;  DB 1;  Length 1701;
  Best Local Similarity  98.7%;  Pred. No. 5.5e-97;
  Matches  389;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1308 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 1367

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1368 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1427

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1428 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1487

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   1488 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1547

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1548 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1607

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
          |||||||||||||||||||||||||||||||||||| |||||||: |||||||||
Db   1608 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 1667

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||||
Db   1668 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1701


RESULT   5
MSP1_PLAF3
ID   MSP1_PLAF3     STANDARD;      PRT;  1682 AA.
AC   P19598; Q25921;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P190).
GN   MSP-1.
OS   Plasmodium falciparum (isolate ro-33 / Ghana).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5834;
RN   [1]
RP   SEQUENCE OF 1-1061 FROM N.A.
RX   MEDLINE=88166657; PubMed=3327688;
RA   Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT   "A naturally occurring gene encoding the major surface antigen
RT   precursor .p190 of Plasmodium falciparum lacks tripeptide repeats.";
RL   EMBO J. 6:4137-4142(1987).
RN   [2]
RP   SEQUENCE OF 1032-1682 FROM N.A.
RX   MEDLINE=95354793; PubMed=7628566;
RA   Tolle R., Bujard H., Cooper J.A.;
RT   "Plasmodium falciparum: variations within the C-terminal region of
RT   merozoite surface antigen-1.";
RL   Exp. Parasitol. 81:47-54(1995).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M35727; AAA29715.1; -.
DR   EMBL; Y00087; CAA68280.1; -.
DR   EMBL; Z35326; CAA84555.1; -.
DR   PIR; S06286; S06286.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1682       MEROZOITE SURFACE PROTEIN 1.
FT   TRANSMEM   1666   1682       MEMBRANE ANCHOR (POTENTIAL).
FT   CARBOHYD    233    233       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    462    462       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    528    528       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    599    599       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    785    785       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    881    881       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    901    901       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    947    947       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1071   1071       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1178   1178       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1569   1569       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1682 AA;  192462 MW;  C82A1E159948CAD6 CRC64;


  Query Match            98.2%;  Score 2025.5;  DB 1;  Length 1682;
  Best Local Similarity  99.0%;  Pred. No. 6.5e-97;
  Matches  390;  Conservative    1;  Mismatches    2;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1290 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 1349

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1350 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1409

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   1410 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1468

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||  |||||||
Db   1469 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFEAIKK 1528

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1529 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPQNSGCFR 1588

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1589 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1648

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||||
Db   1649 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1682


RESULT   6
MSP1_PLAFK
ID   MSP1_PLAFK     STANDARD;      PRT;  1630 AA.
AC   P04932;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P190).
```

```
GN   MSP-1.
OS   Plasmodium falciparum (isolate K1 / Thailand).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5839;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86136024; PubMed=3004972;
RA   Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA   Stunnenberg H., Bujard H.;
RT   "Polymorphism of the precursor for the major surface antigens of
RT   Plasmodium falciparum merozoites: studies at the genetic level.";
RL   EMBO J. 4:3823-3829(1985).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RA   Pan W., Tolle R., Bujard H.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03371; CAA27070.1; -.
DR   PIR; A25120; SAZQK1.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1630       MEROZOITE SURFACE PROTEIN 1.
FT   DOMAIN       67     84       TRIPEPTIDE SG(TP) REPEAT.
FT   TRANSMEM   1614   1630       MEMBRANE ANCHOR.
FT   CARBOHYD     97     97       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    755    755       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    759    759       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    774    774       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    835    835       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    911    911       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    955    955       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1049   1049       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1156   1156       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1165   1165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1436   1436       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1517   1517       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1630 AA;  187289 MW;  ADBDEC3CE0A46322 CRC64;


  Query Match             62.5%;  Score 1289.5;  DB 1;  Length 1630;
  Best Local Similarity   62.3%;  Pred. No. 3.5e-59;
  Matches  243;  Conservative   54;  Mismatches   72;  Indels   21;  Gaps    2;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db    1261 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 1320

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db    1321 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 1380

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db    1381 SEKYKSDLDSIKKYI--------------------NDKQGENEKYLPFLNNIETLYKTVN 1420

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
           :||| ::|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db    1421 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 1480

Qy     246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDE 304
             ::| | ||||:| :|   |::| |::|  | ||||||||||||||:||||||||
Db    1481 NHNNLLTKFLSTGMVFENLAKTVLSNLLDGNLQGMLNISQHQCVKKQCPQNSGCFRHLDE 1540

Qy     305 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1541 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 1600

Qy     365 PLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||
Db    1601 PLFDGIFCSSSNFLGISFLLILMLILYSFI 1630


RESULT   7
MSP1_PLAFW
ID   MSP1_PLAFW     STANDARD;      PRT;  1639 AA.
AC   P04933;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P195).
GN   MSP-1.
OS   Plasmodium falciparum (isolate Wellcome).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5848;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86014355; PubMed=2995820;
RA   Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA   Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA   Freeman R.R.;
RT   "Primary structure of the precursor to the three major surface
RT   antigens of Plasmodium falciparum merozoites.";
RL   Nature 317:270-273(1985).
RN   [2]
RP   REVISIONS.
RA   Holder A.A.;
RL   Submitted (MAR-1991) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X02919; CAA26676.1; -.
DR   PIR; A24594; A24594.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1639       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    116    116       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    268    268       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    764    764       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    768    768       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    783    783       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    844    844       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    920    920       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    964    964       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1058   1058       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD   1165   1165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1174   1174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1445   1445       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1526   1526       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1639 AA;  187618 MW;  2C255B6616C87F6E CRC64;


  Query Match           62.5%;  Score 1289.5;  DB 1;  Length 1639;
  Best Local Similarity   62.3%;  Pred. No. 3.5e-59;
  Matches  243;  Conservative   54;  Mismatches   72;  Indels   21;  Gaps    2;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |: |||:|: ||||||  ||: | :||
Db    1270 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 1329

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db    1330 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 1389

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db    1390 SEKYKSDLDSIKKYI--------------------NDKQGENEKYLPFLNNIETLYKTVN 1429

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
           :||| ::|:|:|:||: :   ||    ||| :|: || | ||:  || :|:|  |  |  |
Db    1430 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 1489

Qy     246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDE 304
             ::| | ||||:| :|   |::| |::|  | ||||||||||||:||||||||||
Db    1490 NHNNLLTKFLSTGMVFENLAKTVLSNLLDGNLQGMLNISQHQCVKKQCPQNSGCFRHLDE 1549

Qy     305 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1550 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDSY 1609

Qy     365 PLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||
Db    1610 PLFDGIFCSSSNFLGISFLLILMLILYSFI 1639


RESULT   8
MSP1_PLAYO
ID   MSP1_PLAYO     STANDARD;      PRT;  1772 AA.
AC   P13828;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (230 KDA).
GN   MSP-1.
OS   Plasmodium berghei yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5862;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90205979; PubMed=2320061;
RA   Lewis A.P.;
RT   "Sequence analysis upstream of the gene encoding the precursor to the
RT   major merozoite surface antigens of Plasmodium yoelii.";
RL   Mol. Biochem. Parasitol. 39:285-288(1990).
RN   [2]
RP   SEQUENCE OF 1093-1772 FROM N.A.
RC   STRAIN=17XL;
RX   MEDLINE=88124889; PubMed=2448778;
RA   Burns J.M. Jr., Daly T.M., Vaidya A.B., Long C.A.;
RT   "The 3' portion of the gene for a Plasmodium yoelii merozoite surface
RT   antigen encodes the epitope recognized by a protective monoclonal
RT   antibody.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:602-606(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03612; AAA29762.1; -.
DR   EMBL; J04668; AAA29702.1; -.
DR   PIR; A28121; A28121.
DR   PIR; A45532; A45532.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   1772       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    406    406       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    646    646       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    829    829       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1018   1018       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1090   1090       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1408   1408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1446   1446       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1541   1541       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1629   1629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1680   1680       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1521   1521       L -> V (IN REF. 2).
SQ   SEQUENCE   1772 AA;  197230 MW;  9A6291658EB0F45D CRC64;


  Query Match           29.8%;  Score 613.5;  DB 1;  Length 1772;
  Best Local Similarity   33.5%;  Pred. No. 1.8e-24;
  Matches  131;  Conservative   90;  Mismatches  143;  Indels   27;  Gaps    7;

Qy       8 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLES 67
           :||| | ||   :| | |   |:|||| : :   |   :: : ||:: :|| ||:||
Db    1401 DILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQKRNDFLEVLSH 1460

Qy      68 DLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLA 127
           :|  || :|:|:|:| : ::||::::|:  : : ||  : :  ||:  :|| :: |::
Db    1461 ELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFNKMVE 1520

Qy     128 KYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKE--SKFLPFLTNIETLYNNLV 185
             |  | ::|: |   :              |:|::  ||   |::| | ::: ||  ::
Db    1521 LYNTQLAAVKEQIATIE-------------AETNDTNKEEKKKYIPILEDLKGLYETVI 1566

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDD 245
           : ::|   |: ::::   || |  :    |     ||:|:| |  | :        |
Db    1567 GQAEEYSEELQNRLDNYKNEKAEFEILTKNLEKYIQIDEKLDEFVEHAE-------NNKH 1619

Qy     246 TKKDMLGKLLSTGLV-QNFPNTIISKLIEGKFQDMLNIS-QHQCV-KKQCPENSGCFRHL 302
             |  |  :||| :     |::|::     |:| :  :| ||  :  |:|:||||
Db    1620 IASIALNNLNKSGLVGEGESKKILAKMLNMDGMDLLGVDPKHVCVDTRDIPKNAGCFRDD 1679

Qy     303 DEREECKCLLNYKQ-EGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKP 361
           :  || :||| ||: ||: |||| |||: |||||  | |   :|  | ||| | | :|
Db    1680 NGTEEWRCLLGYKKGEGNTCVENNNPTCDINNGGCDPTASCQNAESTENSKKIICTCKEP 1739

Qy     362 DSYPLFDGIFCSSSNFLGISFLLILMLILYS 392
             ::|:|||||:|:|:| |||: ||:::
Db    1740 TPNAYYEGVFCSSSSFMGLSILLIITLIVFN 1770


RESULT   9
YL17_CAEEL
ID   YL17_CAEEL     STANDARD;      PRT;  1130 AA.
AC   Q11102;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
```

```
DE   HYPOTHETICAL 131.5 KDA PROTEIN C02F12.7 IN CHROMOSOME X.
GN   C02F12.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Miller N.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: WEAK, TO MYOSINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U41545; AAA83190.1; -.
DR   WormPep; C02F12.7; CE03901.
KW   Hypothetical protein; Coiled coil.
FT   DOMAIN      121    779       COILED COIL (POTENTIAL).
FT   DOMAIN      805   1061       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1130 AA;  131485 MW;  B0FD2EFE3D99FB09 CRC64;


  Query Match             7.4%;  Score 153;  DB 1;  Length 1130;
  Best Local Similarity   24.1%;  Pred. No. 0.46;
  Matches   79;  Conservative   64;  Mismatches  125;  Indels   60;  Gaps   15;

Qy      31 SLKKQIEKNIFTFNLNLNDILNSRLKK--------RKYFLDVLESDLMQFKHISSN--EY 80
           :|:||:|| | :   : | :|:   :|           : |||::: |||    |  ||
Db     596 NLRKQLEKEI-SHTEDRNRLLHENTQKELEAHKETHTETVRVLEAEIDQFKSAFENEQEY 654

Qy      81 IIEDSFKLLNSEQKN-TLLKSYKYIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKV 139
             | | |:   | :| |||   : :|   ||   |  :  :  |:: :| | ::| :|:
Db     655 GKEKSAKIRELEAQNKTLLSEMEKVKHVAENLEAFTSDKDNLLEELESKNK-NIEHLKQE 713

Qy     140 IKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKIDDYLINLKAKI 199
           | :  ||              | | :|:|:   :  :|   ::  ::|:   : :  :
Db     714 IAQLNEKI------------STKETEKDSELEKTIAQLEIDNSSKSDQIEKLHLRVNDML 761

Qy     200 NDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKL--INDDTKKDMLGKLLST 257
           :     |||    : |  ::| |  |        |  |:  ||  : :: :|:
Db     762 DQMGTIKDEL---VKKNEEIKTISAKTAQLLESNTVESETKLASVTEEREKE-------- 810

Qy     258 GLVQNFPNTIISKLIEGKFQDMLNISQ-----HQCVKKQCPENSGCFRHLDE-----REE 307
             :|:|  | ||:    |  ::|  ::       |:|::  | :|    ||:      ||
Db     811 --IQSF-QTQISE----KDNEVLTKAERINELETCLKEREVELTGMRTKLDDMTQQLNEE 863

Qy     308 CKCLL--NYKQEGDKCVENPNPTCNENN 333
             :|  |  ||    ::    | || |
Db     864 TTVVLFDNSIQEK---IDEKEATINEMN 888


RESULT  10
SMC1_YEAST
ID   SMC1_YEAST     STANDARD;      PRT;  1225 AA.
AC   P32908;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CHROMOSOME SEGREGATION PROTEIN SMC1 (DA-BOX PROTEIN SMC1).
GN   SMC1 OR CHL10 OR YFL008W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94103320; PubMed=8276886;
RA   Strunnikov A.V., Larionov V.L., Koshland D.;
RT   "SMC1: an essential yeast gene encoding a putative head-rod-tail
RT   protein is required for nuclear division and defines a new ubiquitous
RT   protein family.";
RL   J. Cell Biol. 123:1635-1648(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=95400292; PubMed=7670463;
RA   Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA   Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA   Yamazaki M., Tashiro H., Eki T.;
RT   "Analysis of the nucleotide sequence of chromosome VI from
RT   Saccharomyces cerevisiae.";
RL   Nat. Genet. 10:261-268(1995).
CC   -!- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC       PART OF A CHROMOSOME CONDENSATION MOTOR.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH SMC2 OR OLIGOMERS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
CC   -!- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC       FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L00602; AAA16595.1; -.
DR   EMBL; D50617; BAA09230.1; -.
DR   EMBL; D31600; BAA06496.1; -.
DR   PIR; A49464; A49464.
DR   PIR; S41804; S41804.
DR   SGD; S0001886; SMC1.
KW   Mitosis; ATP-binding; Coiled coil; Nuclear protein.
FT   NP_BIND      33     40       ATP (POTENTIAL).
FT   DOMAIN      173    489       COILED COIL (POTENTIAL).
FT   DOMAIN      679   1063       COILED COIL (POTENTIAL).
FT   DOMAIN     1057   1061       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN     1137   1164       ALA/ASP-RICH (DA-BOX).
FT   MUTAGEN     173    173       S->L: IN TS MUTANT SMC1-2.
FT   MUTAGEN     458    458       N->D: IN TS MUTANT SMC1-1.
SQ   SEQUENCE   1225 AA;  141279 MW;  B504017AA0ECCA8C CRC64;


  Query Match             6.8%;  Score 139.5;  DB 1;  Length 1225;
  Best Local Similarity   22.9%;  Pred. No. 2.5;
  Matches   89;  Conservative   62;  Mismatches  110;  Indels  127;  Gaps   21;

Qy      36 IEKNIFTFNLNLNDILNSRLKKRKYFLDVLESDLMQFKHISSNEYIIE----------DS 85
           ||| |         : |    |:::| :::  |:   | | :: ::   |         |
Db     329 IEKRI--------ESLQKDLQRQKTYVERFET---QLKVVTRSKEAFEEEIKQSARNYDK 377

Qy      86 FKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEEKE 145
           |||  ::    . ||:|  : |      |:  || |  |: :|   :|    |: |:|| |
Db     378 FKLNEND-----LKTYNCLHE------KYLTEGGSILEEKIAVLNND----KREIQEELE 422

Qy     146 KFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKIDDYLINLKAKINDCNVE 205
           :|           : |  |:       :|           |:|  | :|:  :|: |
Db     423 RFNK----------RADISKRRITEELSITG---------EKLDTQLNDLRVSLNEKNAL 463

Qy     206 KDEAHVKITKL-SDLKAIDD-------------KI-DLFKNHND-------------- 234
           |   :: || ||::: ::               || ||  |  :
Db     464 HTERLHELKKLQSDIESANNQEYDLNFKLRETLVKIDDLSANQRETMKERKLRENIAMLK 523

Qy     235 --FEAIKKLIND--DTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCV-- 288
           |  :| |::|    ||:  |  :|| | :|| : |:          : | ::| :|:
Db     524 RFFPGVKGLVHDLCHPKKEKYGLAVSTILGKNFDSVIV---------ENLTVAQ-ECIAF 573
```

```
Qy    289 -KKQCPENSGCFRHLDERE-ECKCL-------------LNYKQEGDKCVE---NPNPTC 329
          |||    :  |  ||  | |   |              ::|: | :| ::     :  |
Db    574 LKKQ-RAGTASFIPLDTIETELPTLSLPDSQDYILSINAIDYEPEYEKAMQYVCGDSIIC 632

Qy    330 NENNGGCDADAKCTEEDSGSNGKKITCE 357
          |  |      ||  :   |  || :| |
Db    633 NTLN-----IAKDLKWKKGIRGKLVTIE 655


RESULT  11
MYS2_DICDI
ID   MYS2_DICDI     STANDARD;      PRT;  2116 AA.
AC   P08799;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   MYOSIN II HEAVY CHAIN, NON MUSCLE.
GN   MHCA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87092266; PubMed=3540939;
RA   Warrick H.M., de Lozanne A., Leinwand L.A., Spudich J.A.;
RT   "Conserved protein domains in a myosin heavy chain gene from
RT   Dictyostelium discoideum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN   [2]
RP   PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC   STRAIN=AX2;
RX   MEDLINE=90353583; PubMed=2387408;
RA   Lueck-Vielmeter D., Schleicher M., Grabatin B., Wippler J.,
RA   Gerisch G.;
RT   "Replacement of threonine residues by serine and alanine in a
RT   phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RL   FEBS Lett. 269:239-243(1990).
RN   [3]
RP   PHOSPHORYLATION SITES.
RX   MEDLINE=88112226; PubMed=2828113;
RA   Wagle G., Noegel A., Scheel J., Gerisch G.;
RT   "Phosphorylation of threonine residues on cloned fragments of the
RT   Dictyostelium myosin heavy chain.";
RL   FEBS Lett. 227:71-75(1988).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX   MEDLINE=95345066; PubMed=7619795;
RA   Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,
RA   Rayment I.;
RT   "X-ray structures of the myosin motor domain of Dictyostelium
RT   discoideum complexed with MgADP.BeFx and MgADP.AlF4-.";
RL   Biochemistry 34:8960-8972(1995).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX   MEDLINE=95345067; PubMed=7619796;
RA   Smith C.A., Rayment I.;
RT   "X-ray structure of the magnesium(II)-pyrophosphate complex of the
RT   truncated head of Dictyostelium discoideum myosin to 2.7-A
RT   resolution.";
RL   Biochemistry 34:8973-8981(1995).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX   MEDLINE=96206189; PubMed=8611530;
RA   Smith C.A., Rayment I.;
RT   "X-ray structure of the magnesium(II).ADP.vanadate complex of the
RT   Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";
RL   Biochemistry 35:5404-5417(1996).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX   MEDLINE=97452580; PubMed=9305951;
RA   Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT   "X-ray structures of the MgADP, MgATPgammaS, and MgAMPPNP complexes
RT   of the Dictyostelium discoideum myosin motor domain.";
RL   Biochemistry 36:11619-11628(1997).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX   MEDLINE=98070605; PubMed=9405148;
RA   Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
RT   "X-ray crystal structure and solution fluorescence characterization
RT   of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the
RT   Dictyostelium discoideum myosin motor domain.";
RL   J. Mol. Biol. 274:394-407(1997).
CC   -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
CC       ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC   -!- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
CC       INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
CC       LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
CC       (MLC-2).
CC   -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
CC       CORTEX.
CC   -!- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER
CC       SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
CC       THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC   -!- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
CC       ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
CC       POSITION (688).
CC   -!- SIMILARITY: THE GLOBULAR HEAD AA SEQUENCE SHOWS A HIGH DEGREE OF
CC       SIMILARITY WITH THE GLOBULAR HEAD SEQUENCES OF MUSCLE & NONMUSCLE
CC       HEAVY CHAINS. BY CONTRAST THE ROD SEQUENCE IS LESS CONSERVED, BUT
CC       THE PERIODICITIES OF HYDROPHOBIC & CHARGED RESIDUES, WHICH DICTATE
CC       THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14628; AAA33227.1; -.
DR   PIR; A26655; A26655.
DR   PIR; S00250; S00250.
DR   PDB; 1MMA; 03-DEC-97.
DR   PDB; 1MMD; 17-AUG-96.
DR   PDB; 1MMG; 03-DEC-97.
DR   PDB; 1MMN; 03-DEC-97.
DR   PDB; 1MND; 17-AUG-96.
DR   PDB; 1MNE; 17-AUG-96.
DR   PDB; 1VOM; 23-DEC-96.
DR   PDB; 1LVK; 28-JAN-98.
DR   DictyDb; DD01008; mhcA.
DR   InterPro; IPR000048; -.
DR   InterPro; IPR001609; -.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF00063; myosin_head; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
KW   Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
KW   Heptad repeat pattern; Methylation; Alkylation; Phosphorylation.
FT   DOMAIN        1    816       GLOBULAR HEAD (S1).
FT   DOMAIN      817   2116       RODLIKE TAIL (S2 AND LMM DOMAINS).
FT   DOMAIN      817   2116       COILED COIL (POTENTIAL).
FT   NP_BIND     179    186       ATP.
FT   DOMAIN      638    660       ACTIN-BINDING.
FT   DOMAIN      738    752       ACTIN-BINDING.
FT   MOD_RES     130    130       METHYLATION (DI-) (POTENTIAL).
FT   MOD_RES     678    678       ALKYLATION (SH-1).
FT   MOD_RES    1823   1823       PHOSPHORYLATION (BY MHCK).
FT   MOD_RES    1833   1833       PHOSPHORYLATION (BY MHCK).
```

```
FT   MOD_RES     2029   2029       PHOSPHORYLATION (BY MHCK).
SQ   SEQUENCE   2116 AA;  243871 MW;  2FC3770BB1EE56A1 CRC64;


  Query Match             6.6%;  Score 137;  DB 1;  Length 2116;
  Best Local Similarity   21.9%;  Pred. No. 6;
  Matches   75;  Conservative   63;  Mismatches  132;  Indels   72;  Gaps   12;

Qy      22 LKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDV-LESDLMQFKHISSNEY 80
           :| |      :|:|:|  |         :   :| | |  |:| ||    | :    :
Db    1673 IKSLVAEVDEVKEQLEDEI---------LAKDKLVKAKRALEVELEEVRDQLEEEEDSRS 1723

Qy      81 IIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVI 140
            :||| : | :| :: : | |      || : |         ::   |  ||::::|| :
Db    1724 ELEDSKRRLTTEVED-IKKKY---DAEVEQNTKL--------DEAKKKLTDDVDTLKKQL 1771

Qy     141 KEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV------NKIDDYLIN 194
           ::||:|   |         :  ::: ||:   ||  ::    |        | :  | :
Db    1772 EDEKKKLNES---------ERAKKRLESENEDFLAKLDAEVKNRSRAEKDRKKYEKDLKD 1822

Qy     195 LKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKKLINDDTKKDMLGKL 254
           | |:||    | :  :   ||      :|:||  ::  : | |    | :|| : |::
Db    1823 TKYKLNDEAATKTQTEIGAAKL------EDQIDELRSKLEQEQAKATQADKSKKTLEGEI 1876

Qy     255 LSTGLVQNFPNTIISKL------IEGKFQDMLNISQHQCVKKQCPENSGCFRHLDEREEC 308
           :          |  :|      :||: :::      : |::     |       | |:
Db    1877 DNLRAQIEDEGKIKMRLEKEKRALEGELEEL-----RETVEEAEDSKS-------EAEQS 1924

Qy     309 KCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSN 350
           | |:  : |  :              |   : |||   ||: ||
Db    1925 KRLVELELEDAR-----------RNLQKEIDAKEIAEDAKSN 1955


RESULT  12
RBP1_PLAVB
ID   RBP1_PLAVB     STANDARD;      PRT;  2869 AA.
AC   Q00798;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
GN   RBP1.
OS   Plasmodium vivax (strain Belem).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=31273;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92315338; PubMed=1617731;
RA   Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT   "A reticulocyte-binding protein complex of Plasmodium vivax
RT   merozoites.";
RL   Cell 69:1213-1226(1992).
CC   -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC       HUMAN RETICULOCYTE CELLS.
CC   -!- SUBUNIT: HOMODIMER (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M88097; AAA29743.1; -.
DR   HSSP; P36956; 1AM9.
KW   Malaria; Receptor; Signal; Transmembrane.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   2869       RETICULOCYTE BINDING PROTEIN 1.
FT   DOMAIN       18   2807       EXTRACELLULAR.
FT   TRANSMEM   2808   2826       POTENTIAL.
FT   DOMAIN     2827   2869       CYTOPLASMIC.
FT   SITE       1030   1032       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE       2599   2601       CELL ATTACHMENT SITE (POTENTIAL).
SQ   SEQUENCE   2869 AA;  330213 MW;  B9DBE442205EBCFF CRC64;


  Query Match             6.6%;  Score 136.5;  DB 1;  Length 2869;
  Best Local Similarity   21.7%;  Pred. No. 8.8;
  Matches   82;  Conservative   70;  Mismatches  145;  Indels   81;  Gaps   17;

Qy      34 KQIEKNIFTFNLNLNDI-----------LNSRLKKRKYFLDVL-----ESDLMQFKHIS 76
           : |:| :: |: | :||            |  :| : |  :| |     |: | | :| :
Db    2265 RDIKKELYLFHQN-SDISIVEGGVQNMLALYDKLNEEKREMDELYRNISETKLKQMEHST 2323

Qy      77 SNEYIIEDSFKLLNSEQKNTLLKSYKYIK-----------ESVENDIKFAQEGI----SY 121
              : :  | :|     :||:  | :|           | ::| :|:  | :    :
Db    2324 DVFKPMIELHKGMNETNNKSLLEKEKKLKSVNDHMHSMEAEMIKNGLKYTPESVQNINNI 2383

Qy     122 YEKVLAKYKDDLESIKK-------VIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFL 174
           |  : |: |  || | :        :::| |::|       |        :  |||: :
Db    2384 YSVIEAEVK-TLEEIDRDYGDNYQIVEEHKKQFSILIDRTNALMDDIEIFKKENNYNLME 2442

Qy     175 TNIETLYNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDD--------KI 226
           | ||     :::::||:  :  |:      | ::       | ::|  ||         |:
Db    2443 VNTET-----IHRVNDYIEKITNKLVQAKTEYEQI------LENIKQNDDMLQNIFLKKV 2491

Qy     227 DLFKNHNDFEAIKK----LINDDTKKDMLGKLLSTG-----LVQNFPNTIISKLIEGKFQ 277
           : :   || :||   ::||  :::   :||  |      : :|   |: |  |: | :
Db    2492 SIIEY---FENVKKKKESILNDLYEQE---RLLKIGEHLDEIKRNVTETLSSYEIDQKME 2545

Qy     278 DMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQ--EGDKCVENPNPTCNENNGG 335
           |   |::   ||    |       | |||  :   : ||  : |  : :       :  |
Db    2546 MM---SKNLLEKKSKMMNYTSIYEL-EREANEINRDAKQIKDDDTILNSVLEAAIQKRGD 2601

Qy     336 CDADAKCTEEDSGSNGKK 353
           ||        |   |  |
Db    2602 MDAIFSQMSADRNPNEYK 2619


RESULT  13
YIO9_YEAST
ID   YIO9_YEAST     STANDARD;      PRT;  1679 AA.
AC   P40457;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   HYPOTHETICAL 195.1 KDA PROTEIN IN DNA43-UBI1 INTERGENIC REGION.
GN   YIL149C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z38059; CAA86129.1; -.
DR   PIR; S48385; S48385.
DR   SGD; S0001411; MLP2.
```

```
KW   Hypothetical protein.
SQ   SEQUENCE   1679 AA;  195141 MW;  298950CC52202D8F CRC64;


  Query Match             6.6%;  Score 136;  DB 1;  Length 1679;
  Best Local Similarity   22.4%;  Pred. No. 5.2;
  Matches   82;  Conservative   69;  Mismatches  127;  Indels   88;  Gaps   17;

Qy       3 SVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQ---IEKNIFTFNLNLNDILNSRLKKRK 59
           || :: :     |  |  :|| |  :: :|: :   :|  : |    |     : :|::
Db    1079 SVLIEKVDDTAANNGDKDHLK-LVSLFSNLRHERNSLETKLTTCKREL-----AFVKQKN 1132

Qy      60 YFLDVLESDLMQFKHISSNEY----IIEDSFKLLNSE--QKNTLLKSYKYIKESVEN--- 110
            |:   :|| : : :|  ||    :| | || :  |  | | | ::   :::|::|
Db    1133 DSLEKTINDLQRTQTLSEKEYQCSAVIIDEFKDITKEVTQVNILKENNAILQKSLKNVTE 1192

Qy     111 -------DIKFAQEGISYYEKVLAKYKDDLE-SIKKVIKEEKEKFPSSPPTTPPSPAKTD 162
                  :   || ||  :: | : |: :  :  |::  | |           |  : |
Db    1193 KNREIYKQLNDRQEEISRLQRDLIQTKEQVSINSNKILVYESEMEQCKQRYQDLSQQQKD 1252

Qy     163 EQKKESKFLPFLTNIETLYNNLVNKIDDYLINLKAKI-----------NDCNVEKDEAHV 211
            |||:              | |:| |    || |:           |  |  | :||
Db    1253 AQKKD------------IEKLTNEISD----LKGKLSSAENANADLENKFNRLKKQAHE 1295

Qy     212 KI-----------TKLSDLKAIDDKI--DL-FKNHNDFEAIKKLINDDTKKDMLGKLLST 257
            |:           :|::|||| ||:  || |:|        |:|: |||     :| |
Db    1296 KLDASKKQQAALTNELNELKAIKDKLEQDLHFEN------AKVIDLDTKL-KAHELQSE 1347

Qy     258 GLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQE 317
           : ::        |:|    :: :: :   : |    :|  |          |  :| ::|
Db    1348 DVSRDHEKDTYRTLME----EIESLKKELQIFKTANSSSDAFE--------KLKVNMEKE 1395

Qy     318 GDKCVE 323
            |: ::
Db    1396 KDRIID 1401


RESULT  14
ECM_HUMAN
ID   ECM_HUMAN      STANDARD;      PRT;  1228 AA.
AC   Q13201;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ENDOTHELIAL CELL MULTIMERIN PRECURSOR.
GN   ECM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 368-376.
RC   TISSUE=Endothelial cells;
RX   MEDLINE=95355440; PubMed=7629143;
RA   Hayward C.P.M., Hassell J.A., Denomme G.A., Rachubinski R.A.,
RA   Brown C., Kelton J.G.;
RT   "The cDNA sequence of human endothelial cell multimerin. A unique
RT   protein with RGDS, coiled-coil, and epidermal growth factor-like
RT   domains and a carboxyl terminus similar to the globular domain of
RT   complement C1q and collagens type VIII and X.";
RL   J. Biol. Chem. 270:18246-18251(1995).
CC   -!- FUNCTION: CARRIER PROTEIN FOR PLATELET (BUT NOT PLASMA) FACTOR
CC       V/VA. MAY PLAY A ROLE IN THE STORAGE AND STABILIZATION OF FACTOR V
CC       IN PLATELETS.
CC   -!- SUBUNIT: MULTIMERIC. COMPOSED OF VARYING SIZED, DISULFIDE-LINKED
CC       MULTIMERS, THE SMALLEST OF WHICH IS A HOMOTRIMER. PROTEOLYSIS OF
CC       THE PROMULTIMERIN IN THE N-TERMINAL REGION, LEADS TO THE MATURE
CC       P155 FORM THAT IS STORED IN PLATELETS.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED BY ENDOTHELIAL CELLS AND
CC       MEGAKARYOCYTES. STORED IN PLATELET ALPHA GRANULES AND ENDOTHELIAL
CC       CELL WEIBEL-PALADE BODIES, FOLLOWING ACTIVATION OF THESE CELLS, IT
CC       IS RELEASED AND ATTACHED TO MEGAKARYOCYTES, PLATELETS, ENDOTHELIUM
CC       AND SUBENDOTHELIUM OF BLOOD VESSELS. NOT FOUND IN PLASMA. FOUND IN
CC       VASCULAR TISSUES SUCH AS PLACENTA, LUNG, AND LIVER.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- PTM: EXTENSIVELY N-GLYCOSYLATED.
CC   -!- DISEASE: DEFICIENCY IN MULTIMERIN DUE TO PROTEOLYTIC DEGRADATION
CC       WITHIN THE PLATELET ALPHA GRANULES IS ASSOCIATED WITH AN AUTOSOMAL
CC       DOMINANT BLEEDING DISORDER (FACTOR V QUEBEC).
CC   -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U27109; AAC52065.1; -.
DR   MIM; 601456; -.
DR   InterPro; IPR000152; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001073; -.
DR   Pfam; PF00008; EGF; 1.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   PROSITE; PS01113; C1Q; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   Signal; Glycoprotein; EGF-like domain; Coiled coil.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1228       ENDOTHELIAL CELL MULTIMERIN.
FT   DOMAIN    1041   1077       EGF-LIKE.
FT   DOMAIN    1119   1149       C1Q.
FT   DOMAIN     333    365       COILED COIL (POTENTIAL).
FT   DOMAIN     400    430       COILED COIL (POTENTIAL).
FT   DOMAIN     503    523       COILED COIL (POTENTIAL).
FT   DOMAIN     580    650       COILED COIL (POTENTIAL).
FT   DOMAIN     675    726       COILED COIL (POTENTIAL).
FT   DOMAIN     819    869       COILED COIL (POTENTIAL).
FT   DOMAIN     189    192       POLY-SER.
FT   DOMAIN     309    313       POLY-GLN.
FT   SITE       186    188       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID  1045   1056       BY SIMILARITY.
FT   DISULFID  1050   1065       BY SIMILARITY.
FT   DISULFID  1067   1076       BY SIMILARITY.
FT   CARBOHYD    21     21       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    97     97       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   114    114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   120    120       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   136    136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   344    344       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   431    431       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   507    507       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   541    541       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   618    618       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   680    680       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   729    729       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   783    783       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   816    816       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   828    828       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   840    840       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   921    921       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   933    933       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   942    942       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   981    981       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1020   1020       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1075   1075       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1228 AA;  138071 MW;  EAA98B0A17E2C4CD CRC64;


  Query Match             6.5%;  Score 135;  DB 1;  Length 1228;
  Best Local Similarity   22.3%;  Pred. No. 4.2;
  Matches   75;  Conservative   63;  Mismatches  132;  Indels   66;  Gaps   15;
```

```
Qy      29 YRSLK-KQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLESDLMQFKHISSNEYIIEDSFK 87
           |:::|   ::| |   :| :||: |:           ||  : : |   | |:  :   |
Db     330 YQAMKLTLLQKKIDNISLTVNDVRNT--------YSSLEGKVSEDK---SREF--QSLLK 376

Qy      88 LLNSEQKNTLLKSYKYIKES---VENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEEK 144
             | |:  | |::    ::|     :||:   || ::   | ::   :|||| :::|::
Db     377 GLKSKSINVLIRD--IVREQFKIFQNDM---QETVAQLFKTVSSLSEDLESTRQIIQKVN 431

Qy     145 EKFPSSPPTTPPSPAKTDEQKKESKFL------PFLTNIETLYNNLVNKIDDYLINLKAK 198
           |   |                : ||:       | ||:|  | |::||   :  :
Db     432 ESVVSI-------------AAQQKFVLVQENRPTLTDIVELRNHIVNVRQEMTLT---- 473

Qy     199 INDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHND----FEAIKKLINDDTKKDMLGKL 254
             |     |  || | |           | : |:     :|:: | ::    |::  :|
Db     474 ---CEKPIKELEVKQTHLEGA--------LEQEHSRSILYYESLNKTLS--KLKEVHEQL 520

Qy     255 LSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNY 314
           |||  | :  |   :: :     : :: : |: :|||       |  |  :|     |
Db     521 LSTEQVSDQKNAPAAESVSNNVTEYMS-TLHENIKKQSLMMLQMFEDLHIQESKINNLTV 579

Qy     315 KQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSN 350
             | :|  |:    | :    |  | |   :|:  |
Db     580 SLEMEK--ESLRGECEDMLSKCRNDFKFQLKDTEEN 613
```

```
RESULT  15
USO1_YEAST
ID   USO1_YEAST     STANDARD;      PRT;  1790 AA.
AC   P25386;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   INTRACELLULAR PROTEIN TRANSPORT PROTEIN USO1.
GN   USO1 OR INT1 OR YDL058W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=X2180-1A;
RX   MEDLINE=91185402; PubMed=2010462;
RA   Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA   Yamasaki M.;
RT   "A cytoskeleton-related gene, uso1, is required for intracellular
RT   protein transport in Saccharomyces cerevisiae.";
RL   J. Cell Biol. 113:245-260(1991).
RN   [2]
RP   SEQUENCE OF 782-1790 FROM N.A.
RA   Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA   Kendrick K.E.;
RL   Submitted (FEB-1993) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-8 FROM N.A.
RA   Bai Y., Symington L.S.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC       COMPLEX.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC       MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC       ER AND THE GOLGI COMPLEX.
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC       OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC       COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC   -!- SIMILARITY: BELONGS TO THE P115(TAP)/USO1/YBL047C FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54378; CAA38253.1; -.
DR   EMBL; L03188; AAB00143.1; -.
DR   EMBL; U53668; AAB66659.1; -.
DR   PIR; A38455; A38455.
DR   HSSP; P80220; 1DIP.
DR   SGD; S0002216; USO1.
KW   Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil;
KW   Calcium-binding.
FT   DOMAIN        1    724       GLOBULAR HEAD.
FT   DOMAIN      725   1790       COILED COIL (POTENTIAL).
FT   DOMAIN      465    487       CHARGED (HYPER-HYDROPHILIC).
FT   DOMAIN      991   1790       DISPENSABLE FOR THE PROTEIN FUNCTION.
FT   DOMAIN     1172   1786       ASP/GLU-RICH (ACIDIC).
FT   CA_BIND     475    486       POTENTIAL.
FT   CONFLICT    847    847       G -> E (IN REF. 2).
FT   CONFLICT    924    924       E -> K (IN REF. 2).
FT   CONFLICT   1253   1253       V -> I (IN REF. 2).
FT   CONFLICT   1319   1319       I -> V (IN REF. 2).
FT   CONFLICT   1461   1461       N -> S (IN REF. 2).
FT   CONFLICT   1581   1581       G -> S (IN REF. 2).
FT   CONFLICT   1600   1600       I -> V (IN REF. 2).
FT   CONFLICT   1661   1661       R -> S (IN REF. 2).
FT   CONFLICT   1772   1772       D -> DEEDDEE (IN REF. 2).
SQ   SEQUENCE   1790 AA;  206424 MW;  6CE2B216E9FD4818 CRC64;
```

```
  Query Match             6.3%;  Score 130.5;  DB 1;  Length 1790;
  Best Local Similarity   24.3%;  Pred. No. 11;
  Matches   85;  Conservative   50;  Mismatches  106;  Indels  109;  Gaps   18;

Qy       2 ISVTMDNILSGFENEYDVIYLK--PLAGVYRSLKKQIEKNIFTFNLNLNDI--LNSRLKK 57
             ||   | | |  : | : |  |    | |  :::|   |:     ||  ::|  || ::|:
Db    1179 ISQLNDEITS-TQQENESIKKKNDELEGEVKAMKSTSEEQ---SNLKKSEIDALNLQIKE 1234

Qy      58 RKYFLDVLESDLMQ--------------------FKHISSNEYIIEDSFKLLNSEQKNT- 96
             |   :  |: |::                     ||    :|  :||  ||  || ||:
Db    1235 LKKKNETNEASLLESIKSVESETVKIKELQDECNFKEKEVSE--LED--KLKASEDKNSK 1290

Qy      97 ---LLKSYKYIKESVE---NDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEEK------ 144
              | |  : ||| ::    ::|   | |:   |   | : :|  :|| .  ||:
Db    1291 YLELQKESEKIKEELDAKTTELKIQLEKITNLSKAKEKSESELSRLKKTSSEERKNAEEQ 1350

Qy     145 -EKFPSSPPTTPPSPAKTDEQKKESKFL-PFLTNIETLYNNLVNKIDDYLI------NLK 196
            ||  :          |    :|| | |    : |   |:  :| ::| ||       ||
Db    1351 LEKLKNE------IQIKNQAFEKERKLLNEGSSTITQEYSEKINTLEDELIRLQNENELK 1404

Qy     197 AKINDCNVEKDEAHVKIT-------KLSDLKAIDDKIDLFK----------------NHN 233
           ||  | |   :   | ::        | : :|:: |:|  :|                 |
Db    1405 AKEID-NTRSELEKVSLSNDELLEEKQNTIKSLQDEILSYKDKITRNDEKLLSIERDNKR 1463

Qy     234 DFEA-----------------IKKLINDDTK--------KDMLGKLLST 257
           | |:                  :|||  : :|        |:|: || ||
Db    1464 DLESLKEQLRAAQESKAKVEEGLKKLEEESSKEKAELEKSKEMMKKLEST 1513
```

```
Search completed: August  8, 2001, 12:32:29
Job time:  197 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:30:49 ; Search time 43.63 Seconds
(without alignments)
687.893 Million cell updates/sec

Title: US-09-500-376-3
Perfect score: 2052
Sequence: 1 AISVTMDNILSGFENEYDVI.....SNFLGISFLLILMLILYSFI 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2052	100.0	1701	2	A26868	major merozoite su
2	2033	99.1	1701	2	A54498	major merozoite su
3	2006	97.8	1726	1	SAZQGM	major merozoite su
4	1987.5	96.9	651	2	S47282	merozoite surface
5	1987	96.8	1726	2	A45948	major merozoite su
6	1254.5	61.1	400	2	A45545	major merozoite su
7	1254.5	61.1	1639	2	S05603	major merozoite su
8	1133	55.2	1631	1	SAZQK1	major merozoite su
9	1091	53.2	1640	2	A24594	probable major sur
10	826	40.3	1726	2	A39401	merozoite surface
11	815	39.7	1751	2	A45604	major blood-stage
12	593.5	28.9	680	2	A28121	major merozoite su
13	593.5	28.9	1772	2	A45532	major merozoite su
14	569	27.7	1785	2	A45546	major merozoite su
15	168.5	8.2	2401	2	T28676	rhoptry protein -
16	166	8.1	980	2	E71606	hypothetical prote
17	156.5	7.6	1939	2	T18372	repeat organellar
18	155.5	7.6	2166	2	G70163	hypothetical prote
19	153.5	7.5	1169	2	A64505	P115 homolog - Met
20	153.5	7.5	1619	2	T18499	hypothetical prote
21	152	7.4	2269	2	T28677	rhoptry protein -
22	150	7.3	1187	2	T18355	hypothetical prote
23	149.5	7.3	1127	2	T28317	ORF MSV156 hypothe
24	147	7.2	1130	2	T34081	hypothetical prote
25	147	7.2	1679	2	S48385	hypothetical prote
26	146.5	7.1	1250	2	E81339	probable restricti
27	145.5	7.1	1712	2	C71618	hypothetical prote
28	145.5	7.1	2116	2	A26655	myosin heavy chain
29	142	6.9	1156	2	B70356	chromosome assembl
30	140	6.8	880	2	F75103	conserved hypothet
31	140	6.8	1005	2	A64465	hypothetical prote
32	139	6.8	1979	2	C71622	hypothetical prote
33	139	6.8	3394	2	T18501	hypothetical prote
34	138.5	6.7	671	2	H64502	hypothetical prote
35	138	6.7	793	2	E64545	hypothetical prote
36	137	6.7	624	2	PC6003	surface membrane p
37	137	6.7	821	2	S67087	hypothetical prote
38	136.5	6.7	819	2	E70105	P115 protein homol
39	136	6.6	978	2	A70387	conserved hypothet
40	135.5	6.6	3119	2	T18414	protein g377 - mal
41	135	6.6	442	2	T18507	hypothetical prote
42	135	6.6	652	2	B59102	hypothetical prote
43	135	6.6	1079	2	T18356	membrane protein p
44	135	6.6	1313	2	F96673	hypothetical prote
45	135	6.6	1819	2	A71928	cag island protein

ALIGNMENTS

RESULT 1
A26868
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
C;Accession: A26868
R;Tanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.
J. Mol. Biol. 195, 273-287, 1987
A;Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmod
A;Reference number: A26868; MUID:88011243
A;Accession: A26868
A;Molecule type: DNA
A;Residues: 1-1701 <TAN>
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 100.0%; Score 2052; DB 2; Length 1701;
Best Local Similarity 100.0%; Pred. No. 9.1e-101;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1308 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 1367

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1368 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 1427

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1428 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1487

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1488 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1547

Qy    241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1548 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1607

Qy    301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1608 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 1667

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||||
Db   1668 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1701
```

```
RESULT   2
A54498
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (is
C;Species: Plasmodium falciparum
C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C;Accession: A54498
R;Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.;
Mol. Biochem. Parasitol. 27, 291-302, 1988
A;Title: Variation in the precursor to the major merozoite surface antigens of Plasmodiu
A;Reference number: A54498; MUID:88142999
A;Accession: A54498
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1701 <PET>
A;Cross-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

  Query Match             99.1%;  Score 2033;  DB 2;  Length 1701;
  Best Local Similarity   99.2%;  Pred. No. 9.1e-100;
  Matches  391;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1308 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 1367

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    1368 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1427

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1428 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1487

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1488 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1547

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1548 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1607

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1608 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 1667

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db    1668 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1701


RESULT   3
SAZQGM
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
N;Alternate names: 195K glycoprotein
C;Species: Plasmodium falciparum
C;Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
C;Accession: A23386; S06361
R;Weber, J.L.; Leininger, W.M.; Lyon, J.A.
Nucleic Acids Res. 14, 3311-3323, 1986
A;Title: Variation in the gene encoding a major merozoite surface antigen of the human m
A;Reference number: A23386; MUID:86205236
A;Accession: A23386
A;Molecule type: DNA
A;Residues: 1-1104 <WEB1>
A;Cross-references: EMBL:X03831
R;Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.
Nucleic Acids Res. 16, 1206, 1988
A;Title: Merozoite surface protein sequence from the Camp strain of the human malaria pa
A;Reference number: S06361; MUID:88143999
A;Accession: S06361
A;Molecule type: DNA
A;Residues: 1104-1726 <WEB2>
A;Cross-references: EMBL:X03831
C;Comment: The merozoite stages of different strains have strain-specific surface ant
C;Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1726/Product: major merozoite surface antigen #status predicted <MAT>
F;67-87,91-96,100-105,109-120/Region: 3-residue repeats (S-G-T)
F;757-765/Region: 3-residue repeats (T-E-E)
F;133,272,501,567,638,827,839,924,944,990,1016,1114,1221,1613,1658/Binding site: carb

  Query Match             97.8%;  Score 2006;  DB 1;  Length 1726;
  Best Local Similarity   98.0%;  Pred. No. 2.5e-98;
  Matches  386;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1333 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 1392

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    1393 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1452

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1453 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1512

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1513 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 1572

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1573 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1632

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
           |||||||||||||||||| ||||||||||||||||||||||||||||||||: |||||||||
Db    1633 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1692

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db    1693 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   4
S47282
merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71)
C;Species: Plasmodium falciparum
A;Variety: strain RO-71
C;Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 09-Jun-2000
C;Accession: S47282
R;Tolle, R.; Bujard, H.; Cooper, J.A.
submitted to the EMBL Data Library, July 1994
A;Description: Plasmodium falciparum: recombination within the C-terminal region of m
A;Reference number: S47282
A;Accession: S47282
A;Molecule type: DNA
A;Residues: 1-651 <TOL>
A;Cross-references: EMBL:Z35329; NID:g535257; PIDN:CAA84558.1; PID:g535258
A;Experimental source: strain RO-71
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen

  Query Match             96.9%;  Score 1987.5;  DB 2;  Length 651;
  Best Local Similarity   97.5%;  Pred. No. 7.8e-98;
  Matches  384;  Conservative    2;  Mismatches    7;  Indels    1;  Gaps    1;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 318
```

```
Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db     319 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 378

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     379 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 437

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     438 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 497

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||||| |||||||||||||||||||||||||:||||||
Db     498 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPQNSGCFR 557

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
           |||||||||||||||||| |||||||||||||||| ||||||||: |||||||||
Db     558 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 617

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db     618 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 651


RESULT   5
A45948
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
C;Species: Plasmodium falciparum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jun-2000
C;Accession: A45948
R;Chang, S.P.; Kramer, K.J.; Yamaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.
Exp. Parasitol. 67, 1-11, 1988
A;Title: Plasmodium falciparum: gene structure and hydropathy profile of the major meroz
A;Reference number: A45948; MUID:89005525
A;Accession: A45948
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1726 <CHA>
A;Cross-references: GB:M37213
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match             96.8%;  Score 1987;  DB 2;  Length 1726;
  Best Local Similarity   97.2%;  Pred. No. 2.5e-97;
  Matches  383;  Conservative    1;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1333 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 1392

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    1393 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1452

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||||
Db    1453 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTTPSPVKTDEQKKESKFLPFLTNIETL 1512

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           ||||||| |||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    1513 YNNLVNKNDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 1572

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    1573 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1632

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
           |||||||||||||||||| |||||||||||||||| ||||||||: |||||||||
Db    1633 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1692

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db    1693 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   6
A45545
major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments
C;Species: Plasmodium falciparum
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
C;Accession: A45545
R;Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.
Mol. Biochem. Parasitol. 49, 29-33, 1991
A;Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protei
A;Reference number: A45545; MUID:92131048
A;Accession: A45545
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-400 <BLA>
A;Note: sequence extracted from NCBI backbone (NCBIN:77612, NCBIP:77621)
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat


  Query Match             61.1%;  Score 1254.5;  DB 2;  Length 400;
  Best Local Similarity   61.2%;  Pred. No. 2.1e-59;
  Matches  240;  Conservative   54;  Mismatches   73;  Indels   25;  Gaps    3;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |::|||:|: |||||| ||: | :||
Db      31 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 90

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db      91 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 150

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                     |:| :  |:|||| ||||||  :
Db     151 SEKYKSDLDSIKKYI--------------------NDKQGENEKYLPFLNNIETLYKTVN 190

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
           :||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db     191 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 250

Qy     246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHL 302
             ::| | ||||:|  | |   |::| |::|  | ||||||||||||||:||||||||
Db     251 NHNNLLTKFLSTGMV--FENLAKTVLSNLLDGNLQGMLNISQHQCVKKQCPQNSGCFRHL 308

Qy     303 DEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPD 362
           |||||||||||||||||| |||||||||||||||| ||||||||: ||||||||||
Db     309 DEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPD 368

Qy     363 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||
Db     369 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 400


RESULT   7
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N;Alternate names: gp195 surface antigen
C;Species: Plasmodium falciparum
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C;Accession: S05603; S04850
R;Myler, P.J.
submitted to the EMBL Data Library, April 1989
A;Reference number: S05603
A;Accession: S05603
A;Molecule type: mRNA
A;Residues: 1-1639 <MYL>
A;Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
R;Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989
```

```
A;Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plasm
A;Reference number: S04850; MUID:89345116
A;Accession: S04850
A;Molecule type: mRNA
A;Residues: 1504-1639 <MYL2>
A;Cross-references: EMBL:X15063
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1639/Product: major merozoite surface antigen #status predicted <MAT>


  Query Match            61.1%;  Score 1254.5;  DB 2;  Length 1639;
  Best Local Similarity  61.2%;  Pred. No. 1e-58;
  Matches  240;  Conservative   54;  Mismatches   73;  Indels   25;  Gaps    3;

Qy      6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
          :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db   1270 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 1329

Qy     66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
          ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db   1330 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 1389

Qy    126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
            ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db   1390 SEKYKSDLDSIKKYI-------------------NDKQGENEKYLPFLNNIETLYKTVN 1429

Qy    186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
          :||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db   1430 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 1489

Qy    246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHL 302
            ::| | ||||:|  | |   |::| |::|  | ||||||||||||||:||||||||
Db   1490 NHNNLLTKFLSTGMV--FENLAKTVLSNLLDGNLQGMLNISQHQCVKKQCPQNSGCFRHL 1547

Qy    303 DEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPD 362
          ||||||||||||||||| |||||||||||||||| ||||||||: ||||||||||
Db   1548 DEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPD 1607

Qy    363 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          ||||||||||||||||||||||||||||||||
Db   1608 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1639


RESULT   8
SAZQK1
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
C;Species: Plasmodium falciparum
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000
C;Accession: A25120
R;Mackay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.; E
EMBO J. 4, 3823-3829, 1985
A;Title: Polymorphism of the precursor for the major surface antigens of Plasmodium falc
A;Reference number: A91030; MUID:86136024
A;Accession: A25120
A;Molecule type: DNA
A;Residues: 1-1631 <MAC>
C;Comment: The merozoite stages of different strains have strain-specific surface antige
C;Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The me
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane prote
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1631/Product: major merozoite surface antigen #status predicted <MAT>
F;67-84/Region: 3-residue repeats (S-G-T/P)
F;1614-1631/Domain: membrane anchor #status predicted <MBN>
F;97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (Asn) (

  Query Match            55.2%;  Score 1133;  DB 1;  Length 1631;
  Best Local Similarity  57.0%;  Pred. No. 2.5e-52;
  Matches  224;  Conservative   57;  Mismatches   86;  Indels   26;  Gaps    4;

Qy      6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
          :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db   1261 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 1320

Qy     66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
          ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db   1321 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 1380

Qy    126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
            ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db   1381 SEKYKSDLDSIKKYI-------------------NDKQGENEKYLPFLNNIETLYKTVN 1420

Qy    186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
          :||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db   1421 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELIYLKTIQDKLADFKKNNNFVGIADLSTDY 1480

Qy    246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRH 301
            ::| | ||||:|  | |   :|:| |::  |  : : : :     |    :|||||||
Db   1481 NHNNLLTKFLSTGMV--FENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRH 1538

Qy    302 LDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKP 361
          |||||||||||||||| || ||||||||||||||||| ||||||||: ||||||||||
Db   1539 LDEREECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKP 1598

Qy    362 DSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db   1599 DSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1631


RESULT   9
A24594
probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodi
C;Species: Plasmodium falciparum
C;Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jun-2000
C;Accession: A24594
R;Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholl
Nature 317, 270-273, 1985
A;Title: Primary structure of the precursor to the three major surface antigens of Pl
A;Reference number: A24594; MUID:86014355
A;Accession: A24594
A;Molecule type: DNA
A;Residues: 1-1640 <HOL>
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match            53.2%;  Score 1091;  DB 2;  Length 1640;
  Best Local Similarity  55.2%;  Pred. No. 4.2e-50;
  Matches  217;  Conservative   60;  Mismatches   90;  Indels   26;  Gaps    4;

Qy      6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
          :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db   1270 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 1329

Qy     66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
          ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db   1330 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 1389

Qy    126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
            ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db   1390 SEKYKSDLDSIKKYI-------------------NDKQGENEKYLPFLNNIETLYKTVN 1429

Qy    186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
          :||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db   1430 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 1489

Qy    246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKF-QDMLNISQHQCVKKQCPENSGCFRH 301
            ::| | ||||:|  | |   :|:| |::  |  : : : :     |    ::|||||
Db   1490 NHNNLLTKFLSTGMV--FENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRH 1547

Qy    302 LDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKP 361
          |||||||||||||||| || || |||||||||||||| ||||||||: ||||||||||
Db   1548 LDEREECKCLLNYKQEGSKCVENSNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKP 1607
```

```
Qy    362 DSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          | |||   |||||||||||||||||||||||||
Db   1608 DCYPLSMVIFCSSSNFLGISFLLILMLILYSFI 1640
```

RESULT 10
A39401
merozoite surface antigen 1 precursor - Plasmodium vivax
C;Species: Plasmodium vivax
C;Date: 28-Feb-1992 #sequence_revision 28-Feb-1992 #text_change 24-Nov-1999
C;Accession: A39401
R;del Portillo, H.A.; Longacre, S.; Khouri, E.; David, P.H.
Proc. Natl. Acad. Sci. U.S.A. 88, 4030-4034, 1991
A;Title: Primary structure of the merozoite surface antigen 1 of Plasmodium vivax reveal
A;Reference number: A39401; MUID:91219506
A;Accession: A39401
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1726 <DEL>
A;Cross-references: GB:M60807
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

Query Match 40.3%; Score 826; DB 2; Length 1726;
Best Local Similarity 43.2%; Pred. No. 4.2e-36;
Matches 167; Conservative 79; Mismatches 119; Indels 22; Gaps 7;

```
Qy     15 NEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLESDLMQFKH 74
          ::|||:|||||||:|: :|||:| ::  || |: |:|:|||||| |||:|| |||  ||:
Db   1346 SDYDVVYLKPLAGMYKKIKKQLENHVNAFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKY 1405

Qy     75 ISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKVLAKYKDDL- 133
           | ||||:| :|||: |:|  || |||||  |:: |:  | :|::|| |:   ||  |
Db   1406 SPSGEYIIKDPYKLLDLEKKKKLLGSYKYIGASIDKDLGTANDGVNYYNKMGELYKTHLT 1465

Qy    134 ---ESIKKV---IKEEKEKFPSSPPTTPPSPAKTDEQKKES---KFLPFLTNIETLYNNL 184
             | :|||   || | :|       :  :  ||    |::   |:|||| :::  | :|
Db   1466 AVNEEVKKVEADIKAEDDKIKKIGSDSTKTTEKTQSMAKKAELEKYLPFLNSLQKEYESL 1525

Qy    185 VNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLIND 244
          |:|:: |  |||  ||:| :|| || : : || |   :|:|:: :| :       ||  |:
Db   1526 VSKVNTYTDNLKKVINNCQLEKKEAEITVKKLQDYNKMDEKLEEYKKSE-----KK--NE 1578

Qy    245 DTKKDMLGKLLSTGLV-QIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLD 303
             :| ||: : |: :     |:|:|:  : | :   |:| |:    |:|: |:|:||
Db   1579 VKSSGLLEKLMKSKLIKENESKEILSQLLNVQTQLLTMSSEHTCIDTNVPDNAACYRYLD 1638

Qy    304 EREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPDS 363
          || :||| :|:|| ||    | || :|||||  :| |   ||     || |:|||  |
Db   1639 GMEEWRCLLTFKEEGGKCVPGSNVTCKDNNGGCAPEAECKMTDS----NKIVCKCTKEGS 1694

Qy    364 YPLFDGIFCSSSNFLGISFLLILMLIL 390
           |||:|:|||||:|| :|||| ::| |
Db   1695 EPLFEGVFCSSSSFLSLSFLLLMLLFL 1721

RESULT 11
A45604
major blood-stage surface antigen Pv200 - Plasmodium vivax
C;Species: Plasmodium vivax
C;Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
C;Accession: A45604
R;Gibson, H.L.; Tucker, J.E.; Kaslow, D.C.; Krettli, A.U.; Collins, W.E.; Kiefer, M.C.;
Mol. Biochem. Parasitol. 50, 325-333, 1992
A;Title: Structure and expression of the gene for Pv200, a major blood-stage surface ant
A;Reference number: A45604; MUID:92158013
A;Accession: A45604
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1751 <GIB>
A;Cross-references: GB:M75674; NID:g160608; PID:g457336
A;Note: sequence extracted from NCBI backbone (NCBIN:83591, NCBIP:83592)
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

Query Match 39.7%; Score 815; DB 2; Length 1751;
Best Local Similarity 41.0%; Pred. No. 1.6e-35;
Matches 163; Conservative 82; Mismatches 109; Indels 44; Gaps 6;

```
Qy     15 NEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLESDLMQFKH 74
          ::|||:|||||||:|:::|||:| ::  || |: |:|:|||||| |||:|| |||  ||:
Db   1371 SDYDVVYLKPLAGMYKTIKKQLENHVNAFNTNITDMLDSRLKKRNYFLEVLNSDLNPFKY 1430

Qy     75 ISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKVLAKYKDDLE 134
           || ||||:| :|||: |:|  |: ||||  |:: |:  | :|::|| |:   ||  |:
Db   1431 SSSGEYIIKDPYKLLDLEKKKKLIGSYKYIGASIDMDLATANDGVTYYNKMGELYKTHLD 1490

Qy    135 SIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKE----------------------SKFLPF 173
           :|  ||: ::          | ||: |:                          |:|||
Db   1491 GVKTEIKKVEDDI----------KKQDEELKKLGNVNSQDSKKNEFIAKKAELEKYLPF 1539

Qy    174 LTNIETLYNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTN 233
          | :::  | :||:|:: |  ||| ||:| :|| || : : || |   :|:|:: :| :
Db   1540 LNSLQKEYESLVSKVNTYTDNLKKVINNCQLEKKEAEITVKKLQDYNKMDEKLEEYKKSE 1599

Qy    234 DFEAIKKLINDDTKKDMLGKLLSTGLV-QIFPNTIISKLIEGKFQDMLNISQHQCVKKQC 292
             ||  |:     :| ||: : |: :     |:|:|:  : | :   |:| |:
Db   1600 -----KK--NEVKSSGLLEKLMKSKLIKENESKEILSQLLNVQTQLLTMSSEHTCIDTNV 1652

Qy    293 PENSGCFRHLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRK 352
          |:|: |:|:|| || :||| :|:|| ||    | || :|||||  :| |   ||
Db   1653 PDNAACYRYLDGTEEWRCLLTFKEEGGKCVPASNVTCKDNNGGCAPEAECKMTDS----N 1708

Qy    353 KITCECTKPDSYPLFDGIFCSSSNFLGISFLLILMLIL 390
          || |:|||  | |||:|:|||||:|| :|||| ::| |
Db   1709 KIVCKCTKEGSEPLFEGVFCSSSSFLSLSFLLLMLLFL 1746
```

RESULT 12
A28121
major merozoite surface antigen - Plasmodium yoelii (fragment)
C;Species: Plasmodium yoelii
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 26-Aug-1999
C;Accession: A28121
R;Burns Jr., J.M.; Daly, T.M.; Vaidya, A.B.; Long, C.A.
Proc. Natl. Acad. Sci. U.S.A. 85, 602-606, 1988
A;Title: The 3' portion of the gene for a Plasmodium yoelii merozoite surface antigen
A;Reference number: A28121; MUID:88124889
A;Accession: A28121
A;Molecule type: DNA
A;Residues: 1-680 <BUR>
A;Cross-references: GB:J03612; NID:g160678; PID:g160679
A;Experimental source: strain 17XL
A;Note: the authors translated the codon GTA for residue 429 as Leu
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen

Query Match 28.9%; Score 593.5; DB 2; Length 680;
Best Local Similarity 32.5%; Pred. No. 2.7e-24;
Matches 128; Conservative 96; Mismatches 137; Indels 33; Gaps 8;

```
Qy      8 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLES 67
          :||| | ||   :| | |   |:|||| : :   |   :: : ||:: :||  ||:||
Db    309 DILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQKRNDFLEVLSH 368

Qy     68 DLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKVLA 127
          :|  || :|:|:|:| : ::||::::|:  : : ||  : :  ||:   :|| :: |::
Db    369 ELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFNKMVE 428

Qy    128 KYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKE--SKFLPFLTNIETLYNNLV 185
          |   | ::|: |   :           |:|::  ||   |::| | ::: ||  ::
Db    429 VYNTQLAAVKEQIATIE-------------AETNDTNKEEKKKYIPILEDLKGLYETVI 474
```
```

```
Qy    186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLF----KNTNDFEAIKKL 241
          : ::|   |: ::::   || |  :    |     ||:|:| |    :|     :|
Db    475 GQAEEYSEELQNRLDNYKNEKAEFEILTKNLEKYIQIDEKLDEFVEHAENNKHIASI--A 532

Qy    242 INDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNIS-QHQCV-KKQCPENSGCF 299
          :|:  |  ::|:  |          |::|::     |:| :  :| ||  :  |:|:|||
Db    533 LNNLNKSGLVGEGES--------KKILAKMLNMDGMDLLGVDPKHVCVDTRDIPKNAGCF 584

Qy    300 RHLDEREECKCLLNYKQ-EGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCEC 358
          |  :  || :||| ||: ||: | || |||| |||||  |:|   :|  : ||| | |
Db    585 RDDNGTEEWRCLLGYKKGEGNTCVENNNPTCDINNGGCDPTASCQNAESTENSKKIICTC 644

Qy    359 TKPDSYPLFDGIFCSSSNFLGISFLLILMLILYS 392
          :|     ::|:|||||:|:|:| |||: ||:::
Db    645 KEPTPNAYYEGVFCSSSSFMGLSILLIITLIVFN 678


RESULT  13
A45532
major merozoite surface antigen precursor - Plasmodium yoelii
C;Species: Plasmodium yoelii
C;Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 26-Aug-1999
C;Accession: A45532; A45531
R;Lewis, A.P.
Mol. Biochem. Parasitol. 36, 271-282, 1989
A;Title: Cloning and analysis of the gene encoding the 230-kilodalton merozoite surface
A;Reference number: A45532; MUID:90014981
A;Accession: A45532
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1772 <LEW>
A;Cross-references: GB:J04668; NID:g160492; PID:g160493
R;Daly, T.M.; Burns Jr., J.M.; Long, C.A.
Mol. Biochem. Parasitol. 36, 283-285, 1989
A;Title: Precursor to the major merozoite surface antigen of Plasmodium yoelii: cloning
A;Reference number: A45531; MUID:90014982
A;Accession: A45531
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 454-1094 <DAL>
A;Cross-references: GB:J03975; NID:g160081; PID:g160082
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match             28.9%;  Score 593.5;  DB 2;  Length 1772;
  Best Local Similarity   32.5%;  Pred. No. 8e-24;
  Matches  128;  Conservative   96;  Mismatches  137;  Indels   33;  Gaps    8;

Qy      8 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLES 67
          :||| | ||   :| | |   |:|||| : :   |   :: : ||:: :|| ||:||
Db   1401 DILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQKRNDFLEVLSH 1460

Qy     68 DLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKVLA 127
          :|  || :|:|:|:| : ::||::::|:  : : ||  : :  ||:   :|| :: |::
Db   1461 ELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFNKMVE 1520

Qy    128 KYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKE--SKFLPFLTNIETLYNNLV 185
            |   | ::|: |   :              |:|::  ||   |::| | ::: ||  ::
Db   1521 LYNTQLAAVKEQIATIE-------------AETNDTNKEEKKKYIPILEDLKGLYETVI 1566

Qy    186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLF----KNTNDFEAIKKL 241
          : ::|   |: ::::   || |  :    |     ||:|:| |    :|     :|
Db   1567 GQAEEYSEELQNRLDNYKNEKAEFEILTKNLEKYIQIDEKLDEFVEHAENNKHIASI--A 1624

Qy    242 INDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNIS-QHQCV-KKQCPENSGCF 299
          :|:  |  ::|:  |          |::|::     |:| :  :| ||  :  |:|:|||
Db   1625 LNNLNKSGLVGEGES--------KKILAKMLNMDGMDLLGVDPKHVCVDTRDIPKNAGCF 1676

Qy    300 RHLDEREECKCLLNYKQ-EGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCEC 358
          |  :  || :||| ||: ||: | || |||| |||||  |:|   :|  : ||| | |
Db   1677 RDDNGTEEWRCLLGYKKGEGNTCVENNNPTCDINNGGCDPTASCQNAESTENSKKIICTC 1736

Qy    359 TKPDSYPLFDGIFCSSSNFLGISFLLILMLILYS 392
          :|     ::|:|||||:|:|:| |||: ||:::
Db   1737 KEPTPNAYYEGVFCSSSSFMGLSILLIITLIVFN 1770


RESULT  14
A45546
major merozoite surface antigen precursor - Plasmodium chabaudi chabaudi
C;Species: Plasmodium chabaudi chabaudi
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
C;Accession: A45546
R;Deleersnijder, W.; Hendrix, D.; Bendahman, N.; Hanegreefs, J.; Brijs, L.; Hamers-Ca
Mol. Biochem. Parasitol. 43, 231-244, 1990
A;Title: Molecular cloning and sequence analysis of the gene encoding the major meroz
A;Reference number: A45546; MUID:91218805
A;Accession: A45546
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1785 <DEL>
A;Cross-references: GB:M34947; NID:g160597; PID:g160598
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen


  Query Match             27.7%;  Score 569;  DB 2;  Length 1785;
  Best Local Similarity   33.0%;  Pred. No. 1.6e-22;
  Matches  130;  Conservative   92;  Mismatches  138;  Indels   34;  Gaps   11;

Qy      8 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDI---LNSRLKKRKYFLDV 64
          :||  |::| : || | |   |:| || : |    |::   ||   || :|:||  ||||
Db   1415 DILDAFKSENEYIYTKSLGNTYKSFKKHMLKE---FSMIKEDIMTGLNYKLEKRNDFLDV 1471

Qy     65 LESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEK 124
          |  :|  || |::|:::::: ::||::::|:  : : ||  : |  ||: | :|| :: |
Db   1472 LSYELALFKDINTNKFVVKNPYQLLDNDKKDKQMINLKYAIKGVTEDIETATDGIEFFNK 1531

Qy    125 VLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNL 184
          ::  ||  | :: : |        ::  | |     || :||  |: |   ::: ||  :
Db   1532 MIELYKPQLNAVNEQI--------AAIGTEP-----TDAEKK--KYAPIFEDLKGLYETI 1576

Qy    185 VNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLIND 244
          :|  :::   |: |: :  :||    : :  |     ||:|:: |      |: :|  |
Db   1577 LNGAEEFSELLQHKLENYKIEKAGFDILMANLETYIRIDEKLEDF-----VESAEK--NK 1629

Qy    245 DTKKDMLGKLLSTGLV-QIFPNTIISKLIEGKFQDMLNI-SQHQCVKKQCPENSGCFRHL 302
                |  |  :||| :     |::|::     |:| | | | |:    |:|:|||||:
Db   1630 HIASIALNNLNKSGLVTEGESKKILAKMLNMDAMDLLGIGSNHVCISTSTPDNAGCFRYD 1689

Qy    303 DEREECKCLLNYKQ--EGDKCEENPNPTCNENNGGCDADATCTEEDS--GSSRKKITCEC 358
          |  || :||| :|:  :|::|  :  | || |||||| :| | | ::      ||| | |
Db   1690 DGTEEWRCLLGFKKDDDGNRCVADDAPVCNNNNGGCDKNADCREVENTDRDPSKKIVCTC 1749

Qy    359 TKPDSYPLFDGIFCSSSNFLGISFLLILMLILYS 392
          :|:    : |:|||| |:|:| |||: ||:::
Db   1750 KEPNPNAYYAGVFCSSSGFMGLSILLIITLIVFN 1783


RESULT  15
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C;Species: Plasmodium yoelii
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C;Accession: T28676; A45521
R;Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A;Title: Comparison of two members of a multigene family coding for high-molecular ma
A;Reference number: Z20507; MUID:97077455
A;Accession: T28676
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2401 <SIN>
A;Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41263.1
```

R;Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A;Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co
A;Reference number: A45521; MUID:91101660
A;Accession: A45521
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 2260-2401 <KEE>
A;Cross-references: GB:M34281

Query Match 8.2%; Score 168.5; DB 2; Length 2401;
Best Local Similarity 22.6%; Pred. No. 0.29;
Matches 73; Conservative 68; Mismatches 127; Indels 55; Gaps 15;

```
Qy       7 DNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDV-- 64
           : | :  ::::| ||:  :  :|  :   :::: |    |  :||  : |    : ::
Db     285 NEIYNTIKSDFDKIYVGDIEQLYNEMFSVVQESNIEHIENKTEILTLKTKIDNVYNNIQN 344

Qy      65 LESDLMQ--FKHISSNEYIIE---DSFKLLNSEQKNIL---LKSYKYIKESVENDI-KFA 115
           :|:: ::   |:| :|  : |   |  | :  |  | |   |: :|  :: : | | ::|
Db     345 METETVKSHLKNIETNNKLSETILDIIKYIYGEITNELNKTLEDFKNKEKGLSNKIDEYA 404

Qy     116 QEGIS---YYEKVL---AKYKD--DLESIK-KVIKEEKEKFPSSPPTTPPSPAKTDEQKK 166
           :| :    |   :|     | |  ::::|| |  |:  ::|     | ||:  |  :
Db     405 KENVQLNVYKSNILEIKKHYNDQINIDNIKEKEAKQNYDQFKEHMKTIPPNEMKYQKPSI 464

Qy     167 ESKFL--PFLTNIETLYNN-----------------LVNKI-----DDYLINLKAKIND 201
           | | :   ||: :   ||:                  | |||   . |: :   : | ||
Db     465 EIKIMKDEFLSKVNK-YNDFDKVYKEKVESEHNKFTELTNKIKTEVSDEEIKKYENKFND 523

Qy     202 ----CNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKKDMLGKLLST 257
                |  |     :   :: || :|| | :  |||:      : |   |:  |   |:
Db     524 SKSLINETKKSIEEEYQNINTLKKVDDYIKVCLNTNEL-----ITNCHNKQTTLKDKLNQ 578

Qy     258 GLVQIFPNTIISKLIEGKFQDML 280
            :  |     | |:   ||:::|
Db     579 NIKTIKETNSIDKIYTDKFENIL 601
```

Search completed: August 8, 2001, 12:30:54
Job time: 102 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:31:56 ; Search time 60.73 Seconds
(without alignments)
376.342 Million cell updates/sec

Title: US-09-500-376-5
Perfect score: 1982
Sequence: 1 AVTTSVIDNILSKIENEYEV.....SNFLGISFLLILMLILYSFI 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1942	98.0	1654	6	AAP50777	Sequence of the P1
2	1818.5	91.8	1639	19	AAW54145	P. falciparum synt
3	1726.5	87.1	355	20	AAY09372	Merozoite surface
4	1726.5	87.1	355	20	AAY05832	Merozoite surface
5	1726.5	87.1	361	20	AAY09373	Merozoite surface
6	1726.5	87.1	361	20	AAY05833	Merozoite surface
7	1720.5	86.8	376	20	AAY09374	Modified merozoite
8	1720.5	86.8	376	20	AAY05834	Modified merozoite
9	1282	64.7	262	22	AAB37610	Merozoite surface
10	553	27.9	116	18	AAW36103	PfMSP1(p19)A prote
11	553	27.9	116	18	AAW22592	PfMSP1(p19)A prote
12	472	23.8	96	22	AAB37608	Merozoite surface
13	472	23.8	108	22	AAB37609	Merozoite surface
14	456.5	23.0	127	18	AAW22593	PfMSP1(p19)S prote
15	456.5	23.0	127	18	AAW36102	PfMSP1(p19)S prote
16	302	15.2	54	14	AAR41356	MSP1EGF2A EGF2-lik
17	283	14.3	54	14	AAR41357	MSP1EGF2B EGF2-lik
18	243	12.3	93	22	AAB37611	Merozoite surface
19	243	12.3	106	14	AAR41358	P. yoelii combined
20	201	10.1	49	14	AAR41354	MSP1EGF1A EGF1-lik
21	198	10.0	49	14	AAR41355	MSP1EGF1B EGF1-lik
22	186	9.4	350	21	AAY70278	Recombinant vaccin
23	155.5	7.8	3973	21	AAB18253	Plasmodium falcipa
24	154	7.8	2010	21	AAB18218	Plasmodium falcipa
25	149	7.5	507	21	AAB18173	Plasmodium falcipa
26	149	7.5	722	21	AAB18291	Plasmodium falcipa
27	149	7.5	980	21	AAB18294	Plasmodium falcipa
28	149	7.5	1346	21	AAB18236	Plasmodium falcipa
29	145	7.3	1714	21	AAB18275	Plasmodium falcipa
30	142	7.2	1254	11	AAR07503	Merozite apical-en
31	142	7.2	1254	18	AAW24575	Merozoite apical-e
32	141	7.1	2500	21	AAB18272	Plasmodium falcipa
33	140	7.1	1308	21	AAB18167	Plasmodium falcipa
34	138.5	7.0	2013	21	AAB18265	Plasmodium falcipa
35	137	6.9	1121	21	AAB18241	Plasmodium falcipa
36	135	6.8	1979	21	AAB18171	Plasmodium falcipa
37	134	6.8	1150	21	AAY97001	S. cerevisiae esse
38	133	6.7	1712	21	AAB18205	Plasmodium falcipa
39	132	6.7	558	21	AAB18273	Plasmodium falcipa
40	132	6.7	1817	21	AAB18301	Plasmodium falcipa
41	131.5	6.6	469	20	AAY20058	B. burgdorferi ant
42	131	6.6	2539	21	AAB18198	Plasmodium falcipa
43	130	6.6	504	21	AAB18228	Plasmodium falcipa
44	129.5	6.5	481	20	AAY20047	B. burgdorferi ant
45	129.5	6.5	497	20	AAY20046	B. burgdorferi ant

ALIGNMENTS

RESULT 1
AAP50777
ID AAP50777 standard; Protein; 1654 AA.
XX
AC AAP50777;
XX
DT 30-SEP-1991 (first entry)
XX
DE Sequence of the P195 protein of Plasmodium falciparum.
XX
KW Malaria vaccine; epitope; antigen; immunogen.
XX
OS Plasmodium falciparum.
XX
PN EP154454-A.
XX
PD 11-SEP-1985.
XX
PF 21-FEB-1985; 85EP-0301173.
XX
PR 26-SEP-1984; 84GB-0024340.
PR 22-FEB-1984; 84GB-0004692.
PR 21-FEB-1985; 85GB-0004429.
XX
PA (WELL) WELLCOME FOUNDATION LTD.
XX
PI Holder A, Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;
XX
DR WPI; 1985-224845/37.
DR N-PSDB; AAN50530.
XX
PT Cloned DNA sequence encoding plasmodium falciparum protein -
PT useful for expressing the protein for use in vaccines against
PT malaria

```
XX
PS   Claim 6; Fig 1; 51pp; English.
XX
CC   The sequence encoding the P195  protein of Plasmodium falciparum
CC   (AAN50530) and a peptide comprising at least one of its epitopes
CC   (see AAP50777) are claimed. Also claimed is a vaccine for inducing
CC   immunity to malaria comprising the novel peptide or P195 or a
CC   peptide comprising at least one epitope when derived from the new
CC   DNA sequence, together with a carrier.
XX
SQ   Sequence   1654 AA;


  Query Match             98.0%;  Score 1942;  DB 6;  Length 1654;
  Best Local Similarity   98.7%;  Pred. No. 1.7e-130;
  Matches  372;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1278 avttsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 1337

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1338 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 1397

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1398 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 1457

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1458 lnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 1517

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1518 vfenllksvlsnlldwklaryvkhfttpmrkktmiqqssgcfrhldereeckcllnykqe 1577

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
           | ||||| |||||||||||||||||||||||||||||||||||||||||   ||||||||
Db    1578 gskcvensnptcnennggcdadakcteedsgsngkkitcectkpdcyplsmvifcsssnf 1637

Qy     361 LGISFLLILMLILYSFI 377
           |||||||||||||||||
Db    1638 lgisfllilmlilysfi 1654


RESULT   2
AAW54145
ID   AAW54145 standard; Protein; 1639 AA.
XX
AC   AAW54145;
XX
DT   23-SEP-1998  (first entry)
XX
DE   P. falciparum synthetic gp190 protein.
XX
KW   gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW   monoclonal antibody; passive immunisation; parasite.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
PN   WO9814583-A2.
XX
PD   09-APR-1998.
XX
PF   02-OCT-1997;   97WO-EP05441.
XX
PR   02-OCT-1996;   96DE-4040817.
XX
PA   (BUJA/) BUJARD H.
XX
PI   Bujard H,  Pan W,  Tolle R;
XX
DR   WPI; 1998-240088/21.
DR   N-PSDB; AAV21451, AAV35363.
XX
PT   Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT   protein - useful in anti-malaria vaccines, also stabilising genes by
PT   reducing their AT content
XX
PS   Example 1; Fig 3c; 48pp; German.
XX
CC   This sequence represents a modified Plasmodium falciparum gp190/MSP-1
CC   (merozoite surface) protein. The gene encoding this protein has been
CC   stabilised by reducing the AT content of the nucleotide sequence. Such a
CC   protein is useful in vaccines against malaria or for producing monoclonal
CC   antibodies (for passive immunisation). The complete gp190 protein can now
CC   be produced outside the parasite and has, at least over extended regions,
CC   the native pattern of folding. Larger amounts of the protein can be
CC   produced recombinantly than would be possible using the parasites as
CC   source.
XX
SQ   Sequence   1639 AA;


  Query Match             91.8%;  Score 1818.5;  DB 19;  Length 1639;
  Best Local Similarity   93.4%;  Pred. No. 1.2e-121;
  Matches  352;  Conservative    6;  Mismatches   18;  Indels    1;  Gaps    1;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1264 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 1323

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1324 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 1383

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1384 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 1443

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1444 lnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 1503

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
           ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||||
Db    1504 vfenlaktvlsnlldgnl-qgmlnisqhqcvkkqcpqnsgcfrhldereeckcllnykqe 1562

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCSSSNF 360
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    1563 gdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcsssnf 1622

Qy     361 LGISFLLILMLILYSFI 377
           ||| ||||||||||||||
Db    1623 lgiffllilmlilysfi 1639


RESULT   3
AAY09372
ID   AAY09372 standard; Protein; 355 AA.
XX
AC   AAY09372;
XX
DT   31-AUG-1999  (first entry)
XX
DE   Merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal.
XX
OS   Plasmodium falciparum.
XX
PN   WO9920774-A2.
```

```
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22226.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-288313/24.
DR   P-PSDB; AAX56008.
XX
PT   Modified malarial  protein for use in anti-malarial vaccines
XX
PS   Example; Fig 1; 35pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The nucleic acid sequence encoding MSP-1-42
CC   has been modified (see AAX56008) compared to the native sequence (see
CC   AAX56009) such that 306 nucleotide positions have been replaced to
CC   lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
CC   instability motifs while maintaining the same protein amino acid
CC   sequence.  These alterations allow MSP-1-42 to be expressed in
CC   mammalian cell culture and in transgenic mice.  Native MSP-1-12
CC   is known to be difficult to express in cell culture systems,
CC   mammalian cell culture systems or in transgenic animals.  The
CC   invention allows expression of MSP-1 protein in the milk of
CC   transgenic animals, and also provides a DNA vaccine comprising a
CC   vector containing the altered MSP-1-42 sequence.
XX
SQ   Sequence   355 AA;


  Query Match             87.1%;  Score 1726.5;  DB 20;  Length 355;
  Best Local Similarity   93.3%;  Pred. No. 6.1e-116;
  Matches  332;  Conservative    6;  Mismatches   17;  Indels    1;  Gaps    1;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 60

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 120

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 180

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 lnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 240

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
           ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||
Db     241 vfenlaktvlsnlldgnl-qgmlnisqhqcvkkqcpqnsgcfrhldereeckcllnykqe 299

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCS 356
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     300 gdkcvenpnptcnenngqcdadakcteedsgsngkkitcectkpdsyplfdgifcs 355


RESULT   4
AAY05832
ID   AAY05832 standard; Protein; 355 AA.
XX
AC   AAY05832;
XX
DT   02-AUG-1999  (first entry)
XX
DE   Merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal.
XX
OS   Plasmodium falciparum.
XX
PN   WO9920766-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22225.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-302742/25.
DR   N-PSDB; AAX25586.
XX
PT   New modified recombinant nucleic acid sequences useful for producing
PT   malarial DNA vaccine
XX
PS   Disclosure; Fig 1; 43pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The nucleic acid sequence encoding MSP-1-42
CC   has been modified (see AAX25586) compared to the native sequence (see
CC   AAX25587) such that 306 nucleotide positions have been replaced to
CC   lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
CC   instability motifs while maintaining the same protein amino acid
CC   sequence.  These alterations allow MSP-1-42 to be expressed in
CC   mammalian cell culture and in transgenic mice.  The invention
CC   provides modified recombinant nucleic acid sequences and methods for
CC   increasing the mRNA levels and protein expression of proteins that
CC   are difficult to express in cell culture systems, mammalian cell
CC   culture systems or in transgenic animals.  The preferred difficult
CC   protein candidates for expression are those derived from lower
CC   organisms such as parasites, bacteria and viruses that have DNA
CC   coding sequences of high AT content or which have mRNA instability
CC   motifs or rare codons relative to the recombinant expression system
CC   to be used.  The invention allows expression of MSP-1 protein in
CC   the milk of transgenic animals, and also provides a DNA vaccine
CC   comprising a vector containing the altered MSP-1-42 sequence.
XX
SQ   Sequence   355 AA;


  Query Match             87.1%;  Score 1726.5;  DB 20;  Length 355;
  Best Local Similarity   93.3%;  Pred. No. 6.1e-116;
  Matches  332;  Conservative    6;  Mismatches   17;  Indels    1;  Gaps    1;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 60

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 120

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 180

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
```

```
Db    181 lnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 240

Qy    241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
          ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||
Db    241 vfenlaktvlsnlldgnl-qgmlnisqhqcvkkqcpqnsgcfrhldereeckcllnykqe 299

Qy    301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCS 356
          |||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    300 gdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcs 355


RESULT   5
AAY09373
ID   AAY09373 standard; Protein; 361 AA.
XX
AC   AAY09373;
XX
DT   31-AUG-1999  (first entry)
XX
DE   Merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal.
XX
OS   Plasmodium falciparum.
XX
PN   WO9920774-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22226.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-288313/24.
DR   P-PSDB; AAX56009.
XX
PT   Modified malarial  protein for use in anti-malarial vaccines
XX
PS   Example; Fig 2; 35pp; English.
XX
CC   This present sequence comprises a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The C-terminal end of the sequence is
CC   modified to include a 6xHis tag.  A nucleic acid (see AAX56008)
CC   encoding MSP-1-42 has been modified according to a method
CC   of the invention in order to improve expression in mammalian cells
CC   and in transgenic animals by reducing the AT content and removing
CC   mRNA instability motifs.  The invention allows expression of
CC   MSP-1-42 in the milk of transgenic animals, and also provides a DNA
CC   vaccine comprising a vector containing the altered MSP-1-42 nucleic
CC   acid.
XX
SQ   Sequence   361 AA;


  Query Match             87.1%;  Score 1726.5;  DB 20;  Length 361;
  Best Local Similarity   93.3%;  Pred. No. 6.3e-116;
  Matches  332;  Conservative    6;  Mismatches   17;  Indels    1;  Gaps    1;

Qy      1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 60

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 120

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 180

Qy    181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 lnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 240

Qy    241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
          ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||
Db    241 vfenlaktvlsnlldgnl-qgmlnisqhqcvkkqcpqnsgcfrhldereeckcllnykqe 299

Qy    301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCS 356
          |||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    300 gdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcs 355


RESULT   6
AAY05833
ID   AAY05833 standard; Protein; 361 AA.
XX
AC   AAY05833;
XX
DT   02-AUG-1999  (first entry)
XX
DE   Merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal.
XX
OS   Plasmodium falciparum.
XX
PN   WO9920766-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22225.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-302742/25.
DR   N-PSDB; AAX25587.
XX
PT   New modified recombinant nucleic acid sequences useful for producing
PT   malarial DNA vaccine
XX
PS   Disclosure; Fig 2; 43pp; English.
XX
CC   This present sequence comprises a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The C-terminal end of the sequence is
CC   modified to include a 6xHis tag.  Nucleic acids (see AAX25586 and
CC   AAX25593) encoding MSP-1-42 have been modified according to a method
CC   of the invention in order to improve expression in mammalian cells
CC   and in transgenic animals.  The invention provides modified
CC   recombinant nucleic acid sequences and methods for increasing the
CC   mRNA levels and protein expression of proteins that are difficult
CC   to express in cell culture systems, especially mammalian cell
CC   culture systems or in transgenic animals.  The preferred difficult
CC   protein candidates for expression are those derived from lower
CC   organisms such as parasites, bacteria and viruses that have DNA
CC   coding sequences of high AT content or which have mRNA instability
CC   motifs or rare codons relative to the recombinant expression system
```

```
CC   to be used.  The invention allows expression of MSP-1 in the milk
CC   of transgenic animals, and also provides a DNA vaccine comprising a
CC   vector containing the altered MSP-1-42 nucleic acid.
XX
SQ   Sequence   361 AA;


  Query Match             87.1%;  Score 1726.5;  DB 20;  Length 361;
  Best Local Similarity   93.3%;  Pred. No. 6.3e-116;
  Matches  332;  Conservative    6;  Mismatches   17;  Indels    1;  Gaps    1;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 60

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 120

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 180

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 lnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 240

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
           ||||| |:|||||||  | : : : :     |    |:||||||||||||||||||||
Db     241 vfenlaktvlsnlldgnl-qgmlnisqhqcvkkqcpqnsgcfrhldereeckcllnykqe 299

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCS 356
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     300 gdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcs 355


RESULT   7
AAY09374
ID   AAY09374 standard; Protein; 376 AA.
XX
AC   AAY09374;
XX
DT   31-AUG-1999  (first entry)
XX
DE   Modified merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal; mutant.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..15
FT                   /note= "beta-casein signal peptide"
FT   Peptide         371..376
FT                   /note= "6xHis tag"
FT   Misc-difference 197
FT                   /note= "Asn in native MSP-1-42 (N181Q mutation)"
FT   Misc-difference 278
FT                   /note= "Asn in native MSP-1-42 (N262Q mutation)"
XX
PN   WO9920774-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22226.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-288313/24.
DR   P-PSDB; AAX56008.
XX
PT   Modified malarial  protein for use in anti-malarial vaccines
XX
PS   Example; Fig 11; 35pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The sequence has been modified to include
CC   an N-terminal beta-casein signal peptide and an N-terminal 6xHis
CC   tag.  In addition, N181Q and N262Q mutations have been introduced to
CC   eliminate 2 N-glycosylation sites.  These modifications allow the
CC   MSP-1-42 protein to be expressed in the milk of transgenic mice.
XX
SQ   Sequence   376 AA;


  Query Match             86.8%;  Score 1720.5;  DB 20;  Length 376;
  Best Local Similarity   93.0%;  Pred. No. 1.8e-115;
  Matches  331;  Conservative    5;  Mismatches   19;  Indels    1;  Gaps    1;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 75

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 135

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 195

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 lqytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 255

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
           ||||| |:|||||||  |   ::  :     |    |:||||||||||||||||||||
Db     256 vfenlaktvlsnlldgnlqgmlq-isqhqcvkkqcpqnsgcfrhldereeckcllnykqe 314

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCS 356
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     315 gdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcs 370


RESULT   8
AAY05834
ID   AAY05834 standard; Protein; 376 AA.
XX
AC   AAY05834;
XX
DT   02-AUG-1999  (first entry)
XX
DE   Modified merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal; mutant.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..15
FT                   /note= "beta-casein signal peptide"
FT   Peptide         371..376
FT                   /note= "6xHis tag"
```

```
FT   Misc-difference 197
FT                   /note= "Asn in native MSP-1-42 (N181Q mutation)"
FT   Misc-difference 278
FT                   /note= "Asn in native MSP-1-42 (N262Q mutation)"
XX
PN   WO9920766-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22225.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-302742/25.
DR   N-PSDB; AAX25593.
XX
PT   New modified recombinant nucleic acid sequences useful for producing
PT   malarial DNA vaccine
XX
PS   Example; Fig 11; 43pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The sequence has been modified to include
CC   an N-terminal beta-casein signal peptide and an N-terminal 6xHis
CC   tag.  In addition, N181Q and N262Q mutations have been introduced to
CC   eliminate 2 N-glycosylation sites.  These modifications allow the
CC   MSP-1-42 protein to be expressed in the milk of transgenic mice.
CC   The invention generally relates to modified recombinant nucleic
CC   acid sequences and methods for increasing the mRNA levels and
CC   protein expression of proteins that are difficult to express in
CC   cell culture systems, mammalian cell culture systems or in
CC   transgenic animals.  The preferred difficult protein candidates for
CC   expression are those derived from lower organisms such as parasites,
CC   bacteria and viruses that have DNA coding sequences of high AT
CC   content or which have mRNA instability motifs or rare codons
CC   relative to the recombinant expression system to be used.
XX
SQ   Sequence   376 AA;

  Query Match             86.8%;  Score 1720.5;  DB 20;  Length 376;
  Best Local Similarity   93.0%;  Pred. No. 1.8e-115;
  Matches  331;  Conservative    5;  Mismatches   19;  Indels    1;  Gaps    1;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 75

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 135

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 195

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 lqytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 255

Qy     241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQE 300
           ||||| |:|||||||  |   ::  :     |    |:|||||||||||||||||||||
Db     256 vfenlaktvlsnlldgnlqgmlq-isqhqcvkkqcpqnsgcfrhldereeckcllnykqe 314

Qy     301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCS 356
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     315 gdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcs 370


RESULT   9
AAB37610
ID   AAB37610 standard; Protein; 262 AA.
XX
AC   AAB37610;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Merozoite surface protein-133.
XX
KW   Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS   Plasmodium falciparum.
XX
PN   WO200063245-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-GB01558.
XX
PR   20-APR-1999;   99GB-0009072.
PR   13-MAY-1999;   99US-0311817.
PR   25-MAY-1999;   99CA-2271451.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR   WPI; 2001-015762/02.
DR   N-PSDB; AAC68978.
XX
PT   Novel variants of the C-terminal fragment of Plasmodium merozoite
PT   surface protein-1, useful as vaccines for treating or preventing
PT   malaria -
XX
PS   Example 5; Fig 15; 126pp; English.
XX
CC   The present invention relates to non-natural variants of a C-terminal
CC   fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC   non-natural variants have reduced affinity for at least 1 antibody
CC   capable of blocking a second antibody that inhibits the proteolytic
CC   cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC   one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC   MSP-1_4_2, compared to natural MSP-1_1_9. The non-natural variants of the
CC   present invention are useful for immunising a mammal against malaria, and
CC   can be used to treat malaria. The present sequence is MSP-133 protein.
XX
SQ   Sequence   262 AA;

  Query Match             64.7%;  Score 1282;  DB 22;  Length 262;
  Best Local Similarity   97.7%;  Pred. No. 2.6e-84;
  Matches  252;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 AVTTSVIDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRE 60
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 60

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 120

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 180

Qy     181 LNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 240
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 lqytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 240
```

```
Qy    241 VFENLLKSVLSNLLDWKL 258
          ||||| |:||||||| |
Db    241 vfenlaktvlsnlldgnl 258


RESULT 10
AAW36103
ID   AAW36103 standard; Protein; 116 AA.
XX
AC   AAW36103;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)A protein sequence.
XX
KW   Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          1..95
FT                   /note= "amino acids derived from P. falciparum MSP1 p19
FT                        fragment"
FT   Region          96..116
FT                   /note= "glycosylphosphatidylinositol anchoring sequence"
XX
PN   WO9730158-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00290.
XX
PR   14-FEB-1996;   96FR-0001822.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425033/39.
DR   N-PSDB; AAT94550.
XX
PT   Recombinant protein containing the merozoite surface protein-1 p19
PT   fragment - useful in anti-malarial vaccines, diagnosis and protein
PT   purification
XX
PS   Disclosure; Fig 1B; 85pp; French.
XX
CC   This is the amino acid sequence of a recombinant protein comprising
CC   amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
CC   protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
CC   glycosylphosphatidylinositol membrane anchoring sequence.  p19 is the
CC   C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   116 AA;


  Query Match             27.9%;  Score 553;  DB 18;  Length 116;
  Best Local Similarity   93.4%;  Pred. No. 1.3e-32;
  Matches   99;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    272 KTMIQQSSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSG 331
          |    ::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 kkqcpensgcfrhldereeckcllnykqegdkcvenpnptcnennggcdadakcteedsg 70

Qy    332 SNGKKITCECTKPDCYPLFDGIFCSSSNFLGISFLLILMLILYSFI 377
          ||||||||||||| |||||||||||||||||||||||||||||||
Db     71 sngkkitcectkpdsyplfdgifcsssnflgisfllilmlilysfi 116


RESULT 11
AAW22592
ID   AAW22592 standard; Protein; 116 AA.
XX
AC   AAW22592;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)A protein sequence.
XX
KW   Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          1..95
FT                   /note= "amino acids derived from P. falciparum MSP1 p19
FT                        fragment"
FT   Region          96..116
FT                   /note= "glycosylphosphatidylinositol anchoring sequence"
XX
PN   WO9730159-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00291.
XX
PR   14-FEB-1996;   96FR-0001821.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425034/39.
DR   P-PSDB; AAW22592.
XX
PT   Recombinant protein containing Plasmodium merozoite surface
PT   protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT   antibodies for diagnosis and protein purification
XX
PS   Disclosure; Fig 1B; 85pp; French.
XX
CC   This is the amino acid sequence of a recombinant protein comprising
CC   amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
CC   protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
CC   glycosylphosphatidylinositol membrane anchoring sequence.  p19 is the
CC   C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   116 AA;


  Query Match             27.9%;  Score 553;  DB 18;  Length 116;
  Best Local Similarity   93.4%;  Pred. No. 1.3e-32;
  Matches   99;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    272 KTMIQQSSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSG 331
          |    ::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 kkqcpensgcfrhldereeckcllnykqegdkcvenpnptcnennggcdadakcteedsg 70

Qy    332 SNGKKITCECTKPDCYPLFDGIFCSSSNFLGISFLLILMLILYSFI 377
          ||||||||||||| |||||||||||||||||||||||||||||||
Db     71 sngkkitcectkpdsyplfdgifcsssnflgisfllilmlilysfi 116
```

```
RESULT 12
AAB37608
ID   AAB37608 standard; protein; 96 AA.
XX
AC   AAB37608;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Merozoite surface protein-1.
XX
KW   Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS   Plasmodium falciparum.
XX
PN   WO200063245-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-GB01558.
XX
PR   20-APR-1999;   99GB-0009072.
PR   13-MAY-1999;   99US-0311817.
PR   25-MAY-1999;   99CA-2271451.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR   WPI; 2001-015762/02.
XX
PT   Novel variants of the C-terminal fragment of Plasmodium merozoite
PT   surface protein-1, useful as vaccines for treating or preventing
PT   malaria -
XX
PS   Example 2; Page 48; 126pp; English.
XX
CC   The present invention relates to non-natural variants of a C-terminal
CC   fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC   non-natural variants have reduced affinity for at least 1 antibody
CC   capable of blocking a second antibody that inhibits the proteolytic
CC   cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC   one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC   MSP-1_4_2, compared to natural MSP-1_1_9. The present sequence is the
CC   wild-type MSP-1 protein. This sequence was used to generate the variants
CC   of the present invention. The non-natural variants of the present
CC   invention are useful for immunising a mammal against malaria, and can be
CC   used to treat malaria.
XX
SQ   Sequence   96 AA;


  Query Match             23.8%;  Score 472;  DB 22;  Length 96;
  Best Local Similarity   93.2%;  Pred. No. 6.5e-27;
  Matches   82;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy     272 KTMIQQSSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSG 331
           |    |:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9 kkqcpqnsgcfrhldereeckcllnykqegdkcvenpnptcnennggcdadakcteedsg 68

Qy     332 SNGKKITCECTKPDCYPLFDGIFCSSSN 359
           |||||||||||||| ||||||||||||
Db      69 sngkkitcectkpdsyplfdgifcsssn 96


RESULT 13
AAB37609
ID   AAB37609 standard; Protein; 108 AA.
XX
AC   AAB37609;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Merozoite surface protein-119.
XX
KW   Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS   Plasmodium falciparum.
XX
PN   WO200063245-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-GB01558.
XX
PR   20-APR-1999;   99GB-0009072.
PR   13-MAY-1999;   99US-0311817.
PR   25-MAY-1999;   99CA-2271451.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR   WPI; 2001-015762/02.
DR   N-PSDB; AAC68977.
XX
PT   Novel variants of the C-terminal fragment of Plasmodium merozoite
PT   surface protein-1, useful as vaccines for treating or preventing
PT   malaria -
XX
PS   Example 5; Fig 15; 126pp; English.
XX
CC   The present invention relates to non-natural variants of a C-terminal
CC   fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC   non-natural variants have reduced affinity for at least 1 antibody
CC   capable of blocking a second antibody that inhibits the proteolytic
CC   cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC   one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC   MSP-1_4_2, compared to natural MSP-1_1_9. The non-natural variants of the
CC   present invention are useful for immunising a mammal against malaria, and
CC   can be used to treat malaria. The present sequence is MSP-119 protein.
XX
SQ   Sequence   108 AA;


  Query Match             23.8%;  Score 472;  DB 22;  Length 108;
  Best Local Similarity   93.2%;  Pred. No. 7.6e-27;
  Matches   82;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy     272 KTMIQQSSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSG 331
           |    |:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 kkqcpqnsgcfrhldereeckcllnykqegdkcvenpnptcnennggcdadakcteedsg 80

Qy     332 SNGKKITCECTKPDCYPLFDGIFCSSSN 359
           |||||||||||||| ||||||||||||
Db      81 sngkkitcectkpdsyplfdgifcsssn 108


RESULT 14
AAW22593
ID   AAW22593 standard; Protein; 127 AA.
XX
AC   AAW22593;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)S protein sequence.
XX
KW   Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Chimeric - Plasmodium vivax.
OS   Chimeric - Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
```

```
FT                   /note= "signal peptide"
FT   Protein         20..127
FT                   /note= "mature protein"
FT   Region          1..32
FT                   /note= "derived from P. vivax MSP1"
FT   Region          33..34
FT                   /note= "encoded by restriction enzyme sequence used to
FT                          create the chimeric sequence"
FT   Region          35..127
FT                   /note= "derived from P. falciparum C-terminal p19
FT                          fragment of MSP1"
XX
PN   WO9730159-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00291.
XX
PR   14-FEB-1996;   96FR-0001821.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425034/39.
DR   P-PSDB; AAW22592.
XX
PT   Recombinant protein containing Plasmodium merozoite surface
PT   protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT   antibodies for diagnosis and protein purification
XX
PS   Disclosure; Fig 1C; 85pp; French.
XX
CC   This is the amino acid sequence of a chimeric protein comprising amino
CC   acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
CC   linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC   C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC   p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   127 AA;
```

```
  Query Match             23.0%;  Score 456.5;  DB 18;  Length 127;
  Best Local Similarity   72.7%;  Pred. No. 1.2e-25;
  Matches   80;  Conservative   12;  Mismatches   11;  Indels    7;  Gaps    1;

Qy     247 KSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQEGDKCVE 306
           | :::|: :: ::::          |    ::||||||||||||||||||||||||||||
Db      25 kqlvanvdefnisqh-------qcvkkqcpensgcfrhldereeckcllnykqegdkcve 77

Qy     307 NPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCS 356
           |||||||||||||||||||||||||||||||||||| ||||||||||
Db      78 npnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcs 127


RESULT  15
AAW36102
ID   AAW36102 standard; Protein; 127 AA.
XX
AC   AAW36102;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)S protein sequence.
XX
KW   Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
```

```
OS   Chimeric - Plasmodium vivax.
OS   Chimeric - Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /note= "signal peptide"
FT   Protein         20..127
FT                   /note= "mature protein"
FT   Region          1..32
FT                   /note= "derived from P. vivax MSP1"
FT   Region          33..34
FT                   /note= "encoded by restriction enzyme sequence used to
FT                          create the chimeric sequence"
FT   Region          35..127
FT                   /note= "derived from P. falciparum C-terminal p19
FT                          fragment of MSP1"
XX
PN   WO9730158-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00290.
XX
PR   14-FEB-1996;   96FR-0001822.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425033/39.
DR   N-PSDB; AAT94549.
XX
PT   Recombinant protein containing the merozoite surface protein-1 p19
PT   fragment - useful in anti-malarial vaccines, diagnosis and protein
PT   purification
XX
PS   Disclosure; Fig 1C; 85pp; French.
XX
CC   This is the amino acid sequence of a chimeric protein comprising amino
CC   acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
CC   linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC   C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC   p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   127 AA;
```

```
  Query Match             23.0%;  Score 456.5;  DB 18;  Length 127;
  Best Local Similarity   72.7%;  Pred. No. 1.2e-25;
  Matches   80;  Conservative   12;  Mismatches   11;  Indels    7;  Gaps    1;

Qy     247 KSVLSNLLDWKLARYVKHFTTPMRKKTMIQQSSGCFRHLDEREECKCLLNYKQEGDKCVE 306
           | :::|: :: ::::          |    ::||||||||||||||||||||||||||||
Db      25 kqlvanvdefnisqh-------qcvkkqcpensgcfrhldereeckcllnykqegdkcve 77

Qy     307 NPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDCYPLFDGIFCS 356
           |||||||||||||||||||||||||||||||||||| ||||||||||
Db      78 npnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcs 127


Search completed: August  8, 2001, 12:31:57
Job time: 165 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:31:55 ; Search time 60.73 Seconds
(without alignments)
374.345 Million cell updates/sec

Title: US-09-500-376-4
Perfect score: 1967
Sequence: 1 AVTPSVIHNILSKIENEYEV.....SNFLGISFLLILMLILYSFI 375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1901	96.6	1654	6	AAP50777	Sequence of the P1
2	1757.5	89.3	1639	19	AAW54145	P. falciparum synt
3	1665.5	84.7	355	20	AAY09372	Merozoite surface
4	1665.5	84.7	355	20	AAY05832	Merozoite surface
5	1665.5	84.7	361	20	AAY09373	Merozoite surface
6	1665.5	84.7	361	20	AAY05833	Merozoite surface
7	1659.5	84.4	376	20	AAY09374	Modified merozoite
8	1659.5	84.4	376	20	AAY05834	Modified merozoite
9	1246	63.3	262	22	AAB37610	Merozoite surface
10	528	26.8	116	18	AAW36103	PfMSP1(p19)A prote
11	528	26.8	116	18	AAW22592	PfMSP1(p19)A prote
12	447	22.7	96	22	AAB37608	Merozoite surface
13	447	22.7	108	22	AAB37609	Merozoite surface
14	432.5	22.0	127	18	AAW22593	PfMSP1(p19)S prote
15	432.5	22.0	127	18	AAW36102	PfMSP1(p19)S prote
16	278	14.1	54	14	AAR41356	MSP1EGF2A EGF2-lik
17	259	13.2	54	14	AAR41357	MSP1EGF2B EGF2-lik
18	233	11.8	93	22	AAB37611	Merozoite surface
19	232	11.8	106	14	AAR41358	P. yoelii combined
20	200	10.2	49	14	AAR41354	MSP1EGF1A EGF1-lik
21	197	10.0	49	14	AAR41355	MSP1EGF1B EGF1-lik
22	193	9.8	350	21	AAY70278	Recombinant vaccin
23	156	7.9	1714	21	AAB18275	Plasmodium falcipa
24	153	7.8	1346	21	AAB18236	Plasmodium falcipa
25	148	7.5	507	21	AAB18173	Plasmodium falcipa
26	146	7.4	722	21	AAB18291	Plasmodium falcipa
27	145	7.4	980	21	AAB18294	Plasmodium falcipa
28	142.5	7.2	2010	21	AAB18218	Plasmodium falcipa
29	141.5	7.2	3973	21	AAB18253	Plasmodium falcipa
30	139.5	7.1	1817	21	AAB18301	Plasmodium falcipa
31	139.5	7.1	2013	21	AAB18265	Plasmodium falcipa
32	137	7.0	504	21	AAB18228	Plasmodium falcipa
33	137	7.0	558	21	AAB18273	Plasmodium falcipa
34	136.5	6.9	1712	21	AAB18205	Plasmodium falcipa
35	136	6.9	2500	21	AAB18272	Plasmodium falcipa
36	135	6.9	1150	21	AAY97001	S. cerevisiae esse
37	132.5	6.7	1817	21	AAB18255	Plasmodium falcipa
38	132	6.7	1308	21	AAB18167	Plasmodium falcipa
39	131	6.7	469	20	AAY20058	B. burgdorferi ant
40	131	6.7	1182	21	AAB18288	Plasmodium falcipa
41	130.5	6.6	1121	21	AAB18241	Plasmodium falcipa
42	129.5	6.6	2206	21	AAB18254	Plasmodium falcipa
43	129.5	6.6	2485	21	AAB18172	Plasmodium falcipa
44	129	6.6	1254	11	AAR07503	Merozite apical-en
45	129	6.6	1254	18	AAW24575	Merozoite apical-e

ALIGNMENTS

RESULT 1
AAP50777
ID AAP50777 standard; Protein; 1654 AA.
XX
AC AAP50777;
XX
DT 30-SEP-1991 (first entry)
XX
DE Sequence of the P195 protein of Plasmodium falciparum.
XX
KW Malaria vaccine; epitope; antigen; immunogen.
XX
OS Plasmodium falciparum.
XX
PN EP154454-A.
XX
PD 11-SEP-1985.
XX
PF 21-FEB-1985; 85EP-0301173.
XX
PR 26-SEP-1984; 84GB-0024340.
PR 22-FEB-1984; 84GB-0004692.
PR 21-FEB-1985; 85GB-0004429.
XX
PA (WELL) WELLCOME FOUNDATION LTD.
XX
PI Holder A, Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;
XX
DR WPI; 1985-224845/37.
DR N-PSDB; AAN50530.
XX
PT Cloned DNA sequence encoding plasmodium falciparum protein -
PT useful for expressing the protein for use in vaccines against
PT malaria

XX
PS Claim 6; Fig 1; 51pp; English.
XX
CC The sequence encoding the P195 protein of Plasmodium falciparum
CC (AAN50530) and a peptide comprising at least one of its epitopes
CC (see AAP50777) are claimed. Also claimed is a vaccine for inducing
CC immunity to malaria comprising the novel peptide or P195 or a
CC peptide comprising at least one epitope when derived from the new
CC DNA sequence, together with a carrier.
XX
SQ Sequence 1654 AA;

Query Match 96.6%; Score 1901; DB 6; Length 1654;
Best Local Similarity 97.1%; Pred. No. 1.6e-135;
Matches 366; Conservative 3; Mismatches 6; Indels 2; Gaps 1;

```
Qy      1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
          ||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1278 avttsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 1337

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1338 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 1397

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1398 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 1457

Qy    181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
          |||||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1458 lnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 1517

Qy    239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
          ||||||||:||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db   1518 vfenllksvlsnlldwklaryvkhfttpmrkktmiqqssgcfrhldereeckcllnykqe 1577

Qy    299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
          ||||||| |||||||||||||||||||||||||||:||||| |||||||||||||||||||
Db   1578 gskcvensnptcnennggcdadakcteedsgsngkkitcectkpdcyplsmvifcsssnf 1637

Qy    359 LGISFLLILMLILYSFI 375
          |||||||||||||||||
Db   1638 lgisfllilmlilysfi 1654
```

RESULT 2
AAW54145
ID AAW54145 standard; Protein; 1639 AA.
XX
AC AAW54145;
XX
DT 23-SEP-1998 (first entry)
XX
DE P. falciparum synthetic gp190 protein.
XX
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW monoclonal antibody; passive immunisation; parasite.
XX
OS Plasmodium falciparum.
OS Synthetic.
XX
PN WO9814583-A2.
XX
PD 09-APR-1998.
XX
PF 02-OCT-1997; 97WO-EP05441.
XX
PR 02-OCT-1996; 96DE-4040817.
XX
PA (BUJA/) BUJARD H.
XX
PI Bujard H, Pan W, Tolle R;

XX
DR WPI; 1998-240088/21.
DR N-PSDB; AAV21451, AAV35363.
XX
PT Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PT reducing their AT content
XX
PS Example 1; Fig 3c; 48pp; German.
XX
CC This sequence represents a modified Plasmodium falciparum gp190/MSP-1
CC (merozoite surface) protein. The gene encoding this protein has been
CC stabilised by reducing the AT content of the nucleotide sequence. Such a
CC protein is useful in vaccines against malaria or for producing monoclonal
CC antibodies (for passive immunisation). The complete gp190 protein can now
CC be produced outside the parasite and has, at least over extended regions,
CC the native pattern of folding. Larger amounts of the protein can be
CC produced recombinantly than would be possible using the parasites as
CC source.
XX
SQ Sequence 1639 AA;

Query Match 89.3%; Score 1757.5; DB 19; Length 1639;
Best Local Similarity 91.2%; Pred. No. 1.2e-124;
Matches 344; Conservative 7; Mismatches 23; Indels 3; Gaps 2;

```
Qy      1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1264 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 1323

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 1383

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1384 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 1443

Qy    181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
          |||||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1444 lnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 1503

Qy    239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
          ||||| |::|||||  | : : : :     |    ||||||||||||||||||||||||||
Db   1504 vfenlaktvlsnlldgnl-qgmlnisqhqcvkkqcpqnsgcfrhldereeckcllnykqe 1562

Qy    299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCSSSNF 358
          | ||||||||||||||||||||||||||||||||||:||||||||   ||||||||
Db   1563 gdkcvenpnptcnenngcdadakcteedsgsngkkitcectkpdsyplfdgifcsssnf 1622

Qy    359 LGISFLLILMLILYSFI 375
          ||| ||||||||||||
Db   1623 lgiffllilmlilysfi 1639
```

RESULT 3
AAY09372
ID AAY09372 standard; Protein; 355 AA.
XX
AC AAY09372;
XX
DT 31-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX
OS Plasmodium falciparum.
XX
PN WO9920774-A2.

```
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22226.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-288313/24.
DR   P-PSDB; AAX56008.
XX
PT   Modified malarial  protein for use in anti-malarial vaccines
XX
PS   Example; Fig 1; 35pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The nucleic acid sequence encoding MSP-1-42
CC   has been modified (see AAX56008) compared to the native sequence (see
CC   AAX56009) such that 306 nucleotide positions have been replaced to
CC   lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
CC   instability motifs while maintaining the same protein amino acid
CC   sequence.  These alterations allow MSP-1-42 to be expressed in
CC   mammalian cell culture and in transgenic mice.  Native MSP-1-12
CC   is known to be difficult to express in cell culture systems,
CC   mammalian cell culture systems or in transgenic animals.  The
CC   invention allows expression of MSP-1 protein in the milk of
CC   transgenic animals, and also provides a DNA vaccine comprising a
CC   vector containing the altered MSP-1-42 sequence.
XX
SQ   Sequence   355 AA;


  Query Match             84.7%;  Score 1665.5;  DB 20;  Length 355;
  Best Local Similarity   91.0%;  Pred. No. 1.5e-118;
  Matches  324;  Conservative    7;  Mismatches   22;  Indels    3;  Gaps    2;

Qy       1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db       1 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 60

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 120

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 180

Qy     181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
           |||||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||
Db     181 lnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 240

Qy     239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
           ||||| |::|||||  | : : : :     |    ||||||||||||||||||||||||
Db     241 vfenlaktvlsnlldgnl-qgmlnisqhqcvkkqcpqnsgcfrhldereeckcllnykqe 299

Qy     299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCS 354
           | ||||||||||||||||||||||||||||||||||:|||||||||   ||||
Db     300 gdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcs 355


RESULT   4
AAY05832
ID   AAY05832 standard; Protein; 355 AA.
XX
AC   AAY05832;
XX
DT   02-AUG-1999  (first entry)
XX
DE   Merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal.
XX
OS   Plasmodium falciparum.
XX
PN   WO9920766-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22225.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-302742/25.
DR   N-PSDB; AAX25586.
XX
PT   New modified recombinant nucleic acid sequences useful for producing
PT   malarial DNA vaccine
XX
PS   Disclosure; Fig 1; 43pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The nucleic acid sequence encoding MSP-1-42
CC   has been modified (see AAX25586) compared to the native sequence (see
CC   AAX25587) such that 306 nucleotide positions have been replaced to
CC   lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
CC   instability motifs while maintaining the same protein amino acid
CC   sequence.  These alterations allow MSP-1-42 to be expressed in
CC   mammalian cell culture and in transgenic mice.  The invention
CC   provides modified recombinant nucleic acid sequences and methods for
CC   increasing the mRNA levels and protein expression of proteins that
CC   are difficult to express in cell culture systems, mammalian cell
CC   culture systems or in transgenic animals.  The preferred difficult
CC   protein candidates for expression are those derived from lower
CC   organisms such as parasites, bacteria and viruses that have DNA
CC   coding sequences of high AT content or which have mRNA instability
CC   motifs or rare codons relative to the recombinant expression system
CC   to be used.  The invention allows expression of MSP-1 protein in
CC   the milk of transgenic animals, and also provides a DNA vaccine
CC   comprising a vector containing the altered MSP-1-42 sequence.
XX
SQ   Sequence   355 AA;


  Query Match             84.7%;  Score 1665.5;  DB 20;  Length 355;
  Best Local Similarity   91.0%;  Pred. No. 1.5e-118;
  Matches  324;  Conservative    7;  Mismatches   22;  Indels    3;  Gaps    2;

Qy       1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db       1 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 60

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 120

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 180

Qy     181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
```

```
Db    181 lnytyeksnvevkikelnylktiqdkladfknnnfvgiadlstdynhnnlltkflstgm 240

Qy    239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
Db    241 vfenlaktvlsnlldgnl-qgmlnisqhqcvkkqcpqnsgcfrhldereeckcllnykqe 299

Qy    299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCS 354
Db    300 gdkcvenpnptcnenngcdadakcteedsgsngkkitcectkpdsyplfdgifcs 355


RESULT   5
AAY09373
ID   AAY09373 standard; Protein; 361 AA.
XX
AC   AAY09373;
XX
DT   31-AUG-1999  (first entry)
XX
DE   Merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal.
XX
OS   Plasmodium falciparum.
XX
PN   WO9920774-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22226.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-288313/24.
DR   P-PSDB; AAX56009.
XX
PT   Modified malarial  protein for use in anti-malarial vaccines
XX
PS   Example; Fig 2; 35pp; English.
XX
CC   This present sequence comprises a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The C-terminal end of the sequence is
CC   modified to include a 6xHis tag.  A nucleic acid (see AAX56008)
CC   encoding MSP-1-42 has been modified according to a method
CC   of the invention in order to improve expression in mammalian cells
CC   and in transgenic animals by reducing the AT content and removing
CC   mRNA instability motifs.  The invention allows expression of
CC   MSP-1-42 in the milk of transgenic animals, and also provides a DNA
CC   vaccine comprising a vector containing the altered MSP-1-42 nucleic
CC   acid.
XX
SQ   Sequence   361 AA;


  Query Match           84.7%;  Score 1665.5;  DB 20;  Length 361;
  Best Local Similarity   91.0%;  Pred. No. 1.5e-118;
  Matches  324;  Conservative    7;  Mismatches   22;  Indels    3;  Gaps    2;

Qy      1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
Db      1 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 60

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
Db     61 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 120

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
Db    121 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 180

Qy    181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
Db    181 lnytyeksnvevkikelnylktiqdkladfknnnfvgiadlstdynhnnlltkflstgm 240

Qy    239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
Db    241 vfenlaktvlsnlldgnl-qgmlnisqhqcvkkqcpqnsgcfrhldereeckcllnykqe 299

Qy    299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCS 354
Db    300 gdkcvenpnptcnenngcdadakcteedsgsngkkitcectkpdsyplfdgifcs 355


RESULT   6
AAY05833
ID   AAY05833 standard; Protein; 361 AA.
XX
AC   AAY05833;
XX
DT   02-AUG-1999  (first entry)
XX
DE   Merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal.
XX
OS   Plasmodium falciparum.
XX
PN   WO9920766-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22225.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-302742/25.
DR   N-PSDB; AAX25587.
XX
PT   New modified recombinant nucleic acid sequences useful for producing
PT   malarial DNA vaccine
XX
PS   Disclosure; Fig 2; 43pp; English.
XX
CC   This present sequence comprises a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The C-terminal end of the sequence is
CC   modified to include a 6xHis tag.  Nucleic acids (see AAX25586 and
CC   AAX25593) encoding MSP-1-42 have been modified according to a method
CC   of the invention in order to improve expression in mammalian cells
CC   and in transgenic animals.  The invention provides modified
CC   recombinant nucleic acid sequences and methods for increasing the
CC   mRNA levels and protein expression of proteins that are difficult
CC   to express in cell culture systems, especially mammalian cell
CC   culture systems or in transgenic animals.  The preferred difficult
CC   protein candidates for expression are those derived from lower
CC   organisms such as parasites, bacteria and viruses that have DNA
CC   coding sequences of high AT content or which have mRNA instability
CC   motifs or rare codons relative to the recombinant expression system
```

```
CC   to be used.  The invention allows expression of MSP-1 in the milk
CC   of transgenic animals, and also provides a DNA vaccine comprising a
CC   vector containing the altered MSP-1-42 nucleic acid.
XX
SQ   Sequence   361 AA;


  Query Match             84.7%;  Score 1665.5;  DB 20;  Length 361;
  Best Local Similarity   91.0%;  Pred. No. 1.5e-118;
  Matches  324;  Conservative    7;  Mismatches   22;  Indels    3;  Gaps    2;

Qy      1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      1 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 60

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 120

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 180

Qy    181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
          |||||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||
Db    181 lnytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 240

Qy    239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
          ||||| |::|||||  | : : : :     |    ||||||||||||||||||||||||
Db    241 vfenlaktvlsnlldgnl-qgmlnisqhqcvkkqcpqnsgcfrhldereeckcllnykqe 299

Qy    299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCS 354
          | ||||||||||||||||||||||||||||||||||:|||||||||   ||||
Db    300 gdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcs 355


RESULT   7
AAY09374
ID   AAY09374 standard; Protein; 376 AA.
XX
AC   AAY09374;
XX
DT   31-AUG-1999  (first entry)
XX
DE   Modified merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal; mutant.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..15
FT                   /note= "beta-casein signal peptide"
FT   Peptide         371..376
FT                   /note= "6xHis tag"
FT   Misc-difference 197
FT                   /note= "Asn in native MSP-1-42 (N181Q mutation)"
FT   Misc-difference 278
FT                   /note= "Asn in native MSP-1-42 (N262Q mutation)"
XX
PN   WO9920774-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22226.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-288313/24.
DR   P-PSDB; AAX56008.
XX
PT   Modified malarial  protein for use in anti-malarial vaccines
XX
PS   Example; Fig 11; 35pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The sequence has been modified to include
CC   an N-terminal beta-casein signal peptide and an N-terminal 6xHis
CC   tag.  In addition, N181Q and N262Q mutations have been introduced to
CC   eliminate 2 N-glycosylation sites.  These modifications allow the
CC   MSP-1-42 protein to be expressed in the milk of transgenic mice.
XX
SQ   Sequence   376 AA;


  Query Match             84.4%;  Score 1659.5;  DB 20;  Length 376;
  Best Local Similarity   90.7%;  Pred. No. 4.5e-118;
  Matches  323;  Conservative    6;  Mismatches   24;  Indels    3;  Gaps    2;

Qy      1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     16 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 75

Qy     61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 135

Qy    121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 195

Qy    181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
          | ||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||
Db    196 lqytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 255

Qy    239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
          ||||| |::|||||  |   ::  :     |    ||||||||||||||||||||||||
Db    256 vfenlaktvlsnlldgnlqgmlq-isqhqcvkkqcpqnsgcfrhldereeckcllnykqe 314

Qy    299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCS 354
          | ||||||||||||||||||||||||||||||||||:|||||||||   ||||
Db    315 gdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcs 370


RESULT   8
AAY05834
ID   AAY05834 standard; Protein; 376 AA.
XX
AC   AAY05834;
XX
DT   02-AUG-1999  (first entry)
XX
DE   Modified merozoite surface protein MSP-1-42.
XX
KW   MSP-1; merozoite surface protein; malaria; vaccine;
KW   protein engineering; protein expression; codon usage;
KW   transgenic animal; mutant.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..15
FT                   /note= "beta-casein signal peptide"
FT   Peptide         371..376
FT                   /note= "6xHis tag"
```

```
FT   Misc-difference 197
FT                   /note= "Asn in native MSP-1-42 (N181Q mutation)"
FT   Misc-difference 278
FT                   /note= "Asn in native MSP-1-42 (N262Q mutation)"
XX
PN   WO9920766-A2.
XX
PD   29-APR-1999.
XX
PF   20-OCT-1998;   98WO-US22225.
XX
PR   15-MAY-1998;   98US-0085649.
PR   20-OCT-1997;   97US-0062592.
XX
PA   (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI   Chen LH,  Meade H;
XX
DR   WPI; 1999-302742/25.
DR   N-PSDB; AAX25593.
XX
PT   New modified recombinant nucleic acid sequences useful for producing
PT   malarial DNA vaccine
XX
PS   Example; Fig 11; 43pp; English.
XX
CC   The present sequence represents a 42 kDa C-terminal portion of
CC   malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC   important target for the development of a vaccine against
CC   Plasmodium falciparum.  The sequence has been modified to include
CC   an N-terminal beta-casein signal peptide and an N-terminal 6xHis
CC   tag.  In addition, N181Q and N262Q mutations have been introduced to
CC   eliminate 2 N-glycosylation sites.  These modifications allow the
CC   MSP-1-42 protein to be expressed in the milk of transgenic mice.
CC   The invention generally relates to modified recombinant nucleic
CC   acid sequences and methods for increasing the mRNA levels and
CC   protein expression of proteins that are difficult to express in
CC   cell culture systems, mammalian cell culture systems or in
CC   transgenic animals. .The preferred difficult protein candidates for
CC   expression are those derived from lower organisms such as parasites,
CC   bacteria and viruses that have DNA coding sequences of high AT
CC   content or which have mRNA instability motifs or rare codons
CC   relative to the recombinant expression system to be used.
XX
SQ   Sequence   376 AA;

  Query Match             84.4%;  Score 1659.5;  DB 20;  Length 376;
  Best Local Similarity   90.7%;  Pred. No. 4.5e-118;
  Matches  323;  Conservative    6;  Mismatches   24;  Indels    3;  Gaps    2;

Qy       1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      16 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 75

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 135

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 195

Qy     181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
           | |||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||
Db     196 lqytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 255

Qy     239 VFENLLKSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQE 298
           ||||| |::|||||| |   :: :     |   ||||||||||||||||||||||||
Db     256 vfenlaktvlsnlldgnlqgmlq-isqhqcvkkqcpqnsgcfrhldereeckcllnykqe 314

Qy     299 GSKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCS 354
           | ||||||||||||||||||||||||||||||||||:|||||||||   ||||
```

```
Db     315 gdkcvenpnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcs 370


RESULT   9
AAB37610
ID   AAB37610 standard; Protein; 262 AA.
XX
AC   AAB37610;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Merozoite surface protein-133.
XX
KW   Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS   Plasmodium falciparum.
XX
PN   WO200063245-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-GB01558.
XX
PR   20-APR-1999;   99GB-0009072.
PR   13-MAY-1999;   99US-0311817.
PR   25-MAY-1999;   99CA-2271451.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR   WPI; 2001-015762/02.
DR   N-PSDB; AAC68978.
XX
PT   Novel variants of the C-terminal fragment of Plasmodium merozoite
PT   surface protein-1, useful as vaccines for treating or preventing
PT   malaria -
XX
PS   Example 5; Fig 15; 126pp; English.
XX
CC   The present invention relates to non-natural variants of a C-terminal
CC   fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC   non-natural variants have reduced affinity for at least 1 antibody
CC   capable of blocking a second antibody that inhibits the proteolytic
CC   cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC   one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC   MSP-1_4_2, compared to natural MSP-1_1_9. The non-natural variants of the
CC   present invention are useful for immunising a mammal against malaria, and
CC   can be used to treat malaria. The present sequence is MSP-133 protein.
XX
SQ   Sequence   262 AA;

  Query Match             63.3%;  Score 1246;  DB 22;  Length 262;
  Best Local Similarity   95.7%;  Pred. No. 6.3e-87;
  Matches  247;  Conservative    2;  Mismatches    7;  Indels    2;  Gaps    1;

Qy       1 AVTPSVIHNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSPFNKRE 60
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||
Db       1 avtpsvidnilskieneyevlylkplagvyrslkkqlennvmtfnvnvkdilnsrfnkre 60

Qy      61 NFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 nfknvlesdlipykdltssnyvvkdpykflnkekrdkflssynyikdsidtdinfandvl 120

Qy     121 GYYKILSEKYKSDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVIHLEAKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gyykilsekyksdldsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 180

Qy     181 LNYTYEKSNVE--IKELIYLKTIQDKLADFKKNNNFVGIADLSTDYNHNNLLTKFLSTGM 238
           | |||||||||  |||| ||||||||||||||||||||||||||||||||||||||||||
Db     181 lqytyeksnvevkikelnylktiqdkladfkknnnfvgiadlstdynhnnlltkflstgm 240
```

```
Qy    239 VFENLLKSILSNLLDWKL 256
          ||||| |::|||||| |
Db    241 vfenlaktvlsnlldgnl 258


RESULT 10
AAW36103
ID   AAW36103 standard; Protein; 116 AA.
XX
AC   AAW36103;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)A protein sequence.
XX
KW   Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          1..95
FT                   /note= "amino acids derived from P. falciparum MSP1 p19
FT                    fragment"
FT   Region          96..116
FT                   /note= "glycosylphosphatidylinositol anchoring sequence"
XX
PN   WO9730158-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00290.
XX
PR   14-FEB-1996;   96FR-0001822.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425033/39.
DR   N-PSDB; AAT94550.
XX
PT   Recombinant protein containing the merozoite surface protein-1 p19
PT   fragment - useful in anti-malarial vaccines, diagnosis and protein
PT   purification
XX
PS   Disclosure; Fig 1B; 85pp; French.
XX
CC   This is the amino acid sequence of a recombinant protein comprising
CC   amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
CC   protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
CC   glycosylphosphatidylinositol membrane anchoring sequence.  p19 is the
CC   C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   116 AA;


  Query Match           26.8%;  Score 528;  DB 18;  Length 116;
  Best Local Similarity   90.6%;  Pred. No. 6.1e-33;
  Matches   96;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy    270 KTMIQQNSGCFRHLDEREECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSG 329
          |    :||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db     11 kkqcpensgcfrhldereeckcllnykqegdkcvenpnptcnennggcdadakcteedsg 70

Qy    330 SNGKKITCQCTKPDSYPLSMVIFCSSSNFLGISFLLILMLILYSFI 375
          ||||||||:|||||||||   |||||||||||||||||||||||||
Db     71 sngkkitcectkpdsyplfdgifcsssnflgisfllilmlilysfi 116


RESULT 11
AAW22592
ID   AAW22592 standard; Protein; 116 AA.
XX
AC   AAW22592;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)A protein sequence.
XX
KW   Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          1..95
FT                   /note= "amino acids derived from P. falciparum MSP1 p19
FT                    fragment"
FT   Region          96..116
FT                   /note= "glycosylphosphatidylinositol anchoring sequence"
XX
PN   WO9730159-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00291.
XX
PR   14-FEB-1996;   96FR-0001821.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425034/39.
DR   P-PSDB; AAW22592.
XX
PT   Recombinant protein containing Plasmodium merozoite surface
PT   protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT   antibodies for diagnosis and protein purification
XX
PS   Disclosure; Fig 1B; 85pp; French.
XX
CC   This is the amino acid sequence of a recombinant protein comprising
CC   amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
CC   protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
CC   glycosylphosphatidylinositol membrane anchoring sequence.  p19 is the
CC   C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   116 AA;


  Query Match           26.8%;  Score 528;  DB 18;  Length 116;
  Best Local Similarity   90.6%;  Pred. No. 6.1e-33;
  Matches   96;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy    270 KTMIQQNSGCFRHLDEREECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSG 329
          |    :||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db     11 kkqcpensgcfrhldereeckcllnykqegdkcvenpnptcnennggcdadakcteedsg 70

Qy    330 SNGKKITCQCTKPDSYPLSMVIFCSSSNFLGISFLLILMLILYSFI 375
          ||||||||:|||||||||   |||||||||||||||||||||||||
Db     71 sngkkitcectkpdsyplfdgifcsssnflgisfllilmlilysfi 116
```

```
RESULT 12
AAB37608
ID   AAB37608 standard; protein; 96 AA.
XX
AC   AAB37608;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Merozoite surface protein-1.
XX
KW   Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS   Plasmodium falciparum.
XX
PN   WO200063245-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-GB01558.
XX
PR   20-APR-1999;   99GB-0009072.
PR   13-MAY-1999;   99US-0311817.
PR   25-MAY-1999;   99CA-2271451.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR   WPI; 2001-015762/02.
XX
PT   Novel variants of the C-terminal fragment of Plasmodium merozoite
PT   surface protein-1, useful as vaccines for treating or preventing
PT   malaria -
XX
PS   Example 2; Page 48; 126pp; English.
XX
CC   The present invention relates to non-natural variants of a C-terminal
CC   fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC   non-natural variants have reduced affinity for at least 1 antibody
CC   capable of blocking a second antibody that inhibits the proteolytic
CC   cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC   one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC   MSP-1_4_2, compared to natural MSP-1_1_9. The present sequence is the
CC   wild-type MSP-1 protein. This sequence was used to generate the variants
CC   of the present invention. The non-natural variants of the present
CC   invention are useful for immunising a mammal against malaria, and can be
CC   used to treat malaria.
XX
SQ   Sequence   96 AA;


  Query Match             22.7%;  Score 447;  DB 22;  Length 96;
  Best Local Similarity   89.8%;  Pred. No. 6.6e-27;
  Matches   79;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy    270 KTMIQQNSGCFRHLDEREECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSG 329
          |    |||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      9 kkqcpqnsgcfrhldereeckcllnykqegdkcvenpnptcnenngcdadakcteedsg 68

Qy    330 SNGKKITCQCTKPDSYPLSMVIFCSSSN 357
          ||||||||:|||||||||   |||||||
Db     69 sngkkitcectkpdsyplfdgifcsssn 96


RESULT 13
AAB37609
ID   AAB37609 standard; Protein; 108 AA.
XX
AC   AAB37609;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Merozoite surface protein-119.
XX
KW   Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS   Plasmodium falciparum.
XX
PN   WO200063245-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-GB01558.
XX
PR   20-APR-1999;   99GB-0009072.
PR   13-MAY-1999;   99US-0311817.
PR   25-MAY-1999;   99CA-2271451.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR   WPI; 2001-015762/02.
DR   N-PSDB; AAC68977.
XX
PT   Novel variants of the C-terminal fragment of Plasmodium merozoite
PT   surface protein-1, useful as vaccines for treating or preventing
PT   malaria -
XX
PS   Example 5; Fig 15; 126pp; English.
XX
CC   The present invention relates to non-natural variants of a C-terminal
CC   fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC   non-natural variants have reduced affinity for at least 1 antibody
CC   capable of blocking a second antibody that inhibits the proteolytic
CC   cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC   one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC   MSP-1_4_2, compared to natural MSP-1_1_9. The non-natural variants of the
CC   present invention are useful for immunising a mammal against malaria, and
CC   can be used to treat malaria. The present sequence is MSP-119 protein.
XX
SQ   Sequence   108 AA;


  Query Match             22.7%;  Score 447;  DB 22;  Length 108;
  Best Local Similarity   89.8%;  Pred. No. 7.7e-27;
  Matches   79;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy    270 KTMIQQNSGCFRHLDEREECKCLLNYKQEGSKCVENPNPTCNENNGGCDADAKCTEEDSG 329
          |    |||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db     21 kkqcpqnsgcfrhldereeckcllnykqegdkcvenpnptcnenngcdadakcteedsg 80

Qy    330 SNGKKITCQCTKPDSYPLSMVIFCSSSN 357
          ||||||||:|||||||||   |||||||
Db     81 sngkkitcectkpdsyplfdgifcsssn 108


RESULT 14
AAW22593
ID   AAW22593 standard; Protein; 127 AA.
XX
AC   AAW22593;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)S protein sequence.
XX
KW   Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Chimeric - Plasmodium vivax.
OS   Chimeric - Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
```

```
FT                   /note= "signal peptide"
FT   Protein         20..127
FT                   /note= "mature protein"
FT   Region          1..32
FT                   /note= "derived from P. vivax MSP1"
FT   Region          33..34
FT                   /note= "encoded by restriction enzyme sequence used to
FT                          create the chimeric sequence"
FT   Region          35..127
FT                   /note= "derived from P. falciparum C-terminal p19
FT                          fragment of MSP1"
XX
PN   WO9730159-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00291.
XX
PR   14-FEB-1996;   96FR-0001821.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425034/39.
DR   P-PSDB; AAW22592.
XX
PT   Recombinant protein containing Plasmodium merozoite surface
PT   protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT   antibodies for diagnosis and protein purification
XX
PS   Disclosure; Fig 1C; 85pp; French.
XX
CC   This is the amino acid sequence of a chimeric protein comprising amino
CC   acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
CC   linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC   C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC   p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   127 AA;


  Query Match             22.0%;  Score 432.5;  DB 18;  Length 127;
  Best Local Similarity   70.0%;  Pred. No. 1.2e-25;
  Matches   77;  Conservative   12;  Mismatches   14;  Indels    7;  Gaps    1;

Qy     245 KSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQEGSKCVE 304
           | :::|: :: ::::          |    :||||||||||||||||||||||| ||||
Db      25 kqlvanvdefnisqh------qcvkkqcpensgcfrhldereeckcllnykqegdkcve 77

Qy     305 NPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCS 354
           |||||||||||||||||||||||||||||:|||||||||   ||||
Db      78 npnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcs 127


RESULT  15
AAW36102
ID   AAW36102 standard; Protein; 127 AA.
XX
AC   AAW36102;
XX
DT   25-MAR-1998  (first entry)
XX
DE   PfMSP1(p19)S protein sequence.
XX
KW   Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW   Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS   Chimeric - Plasmodium vivax.
OS   Chimeric - Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /note= "signal peptide"
FT   Protein         20..127
FT                   /note= "mature protein"
FT   Region          1..32
FT                   /note= "derived from P. vivax MSP1"
FT   Region          33..34
FT                   /note= "encoded by restriction enzyme sequence used to
FT                          create the chimeric sequence"
FT   Region          35..127
FT                   /note= "derived from P. falciparum C-terminal p19
FT                          fragment of MSP1"
XX
PN   WO9730158-A2.
XX
PD   21-AUG-1997.
XX
PF   14-FEB-1997;   97WO-FR00290.
XX
PR   14-FEB-1996;   96FR-0001822.
XX
PA   (INSP ) INST PASTEUR.
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Barnwell JW,  Longacre-Andre S,  Mendis K,  Nato F;
PI   Roth C;
XX
DR   WPI; 1997-425033/39.
DR   N-PSDB; AAT94549.
XX
PT   Recombinant protein containing the merozoite surface protein-1 p19
PT   fragment - useful in anti-malarial vaccines, diagnosis and protein
PT   purification
XX
PS   Disclosure; Fig 1C; 85pp; French.
XX
CC   This is the amino acid sequence of a chimeric protein comprising amino
CC   acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
CC   linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC   C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC   p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC   The recombinant protein can be used for the production of anti-malarial
CC   vaccines, where the p19 fragment provides a high level of protective
CC   immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ   Sequence   127 AA;


  Query Match             22.0%;  Score 432.5;  DB 18;  Length 127;
  Best Local Similarity   70.0%;  Pred. No. 1.2e-25;
  Matches   77;  Conservative   12;  Mismatches   14;  Indels    7;  Gaps    1;

Qy     245 KSILSNLLDWKLARYVKHFTTPMRKKTMIQQNSGCFRHLDEREECKCLLNYKQEGSKCVE 304
           | :::|: :: ::::          |    :||||||||||||||||||||||| ||||
Db      25 kqlvanvdefnisqh-------qcvkkqcpensgcfrhldereeckcllnykqegdkcve 77

Qy     305 NPNPTCNENNGGCDADAKCTEEDSGSNGKKITCQCTKPDSYPLSMVIFCS 354
           |||||||||||||||||||||||||||||:|||||||||   ||||
Db      78 npnptcnennggcdadakcteedsgsngkkitcectkpdsyplfdgifcs 127


Search completed: August  8, 2001, 12:31:56
Job time: 164 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:12 ; Search time 70.07 Seconds
(without alignments)
743.944 Million cell updates/sec

Title: US-09-500-376-2
Perfect score: 2062
Sequence: 1 AISVTMDNILSGFENEYDVI.....SNFLGISFLLILMLILYSFI 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2043.5	99.1	651	5	Q25924	Q25924 plasmodium
2	2036.5	98.8	569	5	Q25983	Q25983 plasmodium
3	2030.5	98.5	569	5	Q25969	Q25969 plasmodium
4	2030.5	98.5	569	5	Q25974	Q25974 plasmodium
5	2030.5	98.5	569	5	Q25975	Q25975 plasmodium
6	2030.5	98.5	569	5	Q25977	Q25977 plasmodium
7	2030.5	98.5	569	5	Q25979	Q25979 plasmodium
8	2030.5	98.5	1694	5	Q9TZT5	Q9tzt5 plasmodium
9	2030.5	98.5	1694	5	Q9NHX1	Q9nhx1 plasmodium
10	2030.5	98.5	1704	5	Q9TZT4	Q9tzt4 plasmodium
11	2027	98.3	570	5	Q25968	Q25968 plasmodium
12	2027	98.3	570	5	Q9TYG2	Q9tyg2 plasmodium
13	2025	98.2	1720	5	Q25922	Q25922 plasmodium
14	2021	98.0	652	5	Q25923	Q25923 plasmodium
15	2017.5	97.8	569	5	Q25978	Q25978 plasmodium
16	2011.5	97.6	569	5	Q25967	Q25967 plasmodium
17	2011.5	97.6	569	5	Q25970	Q25970 plasmodium
18	2011.5	97.6	569	5	Q25980	Q25980 plasmodium
19	2011.5	97.6	569	5	Q25982	Q25982 plasmodium
20	1961	95.1	373	5	Q25724	Q25724 plasmodium
21	1956	94.9	373	5	Q25721	Q25721 plasmodium
22	1948	94.5	373	5	O43995	O43995 plasmodium
23	1943.5	94.3	372	5	Q25725	Q25725 plasmodium
24	1943	94.2	373	5	O43996	O43996 plasmodium
25	1943	94.2	373	5	Q25723	Q25723 plasmodium
26	1938.5	94.0	372	5	Q25726	Q25726 plasmodium
27	1937.5	94.0	372	5	O43997	O43997 plasmodium
28	1936	93.9	373	5	Q25722	Q25722 plasmodium
29	1924	93.3	373	5	Q25727	Q25727 plasmodium
30	1919	93.1	373	5	Q25728	Q25728 plasmodium
31	1913.5	92.8	372	5	Q25717	Q25717 plasmodium
32	1908.5	92.6	372	5	Q25718	Q25718 plasmodium
33	1908.5	92.6	372	5	Q25719	Q25719 plasmodium
34	1908.5	92.6	372	5	Q25720	Q25720 plasmodium
35	1292.5	62.7	539	5	Q25972	Q25972 plasmodium
36	1292.5	62.7	539	5	Q25981	Q25981 plasmodium
37	1289.5	62.5	400	5	Q03999	Q03999 plasmodium
38	1289.5	62.5	539	5	Q25966	Q25966 plasmodium
39	1289.5	62.5	539	5	Q25976	Q25976 plasmodium
40	1289.5	62.5	539	5	Q25984	Q25984 plasmodium
41	1289.5	62.5	539	5	Q9TYG1	Q9tyg1 plasmodium
42	1283.5	62.2	539	5	Q25971	Q25971 plasmodium
43	1283.5	62.2	539	5	Q25973	Q25973 plasmodium
44	1149	55.7	218	5	Q9TVG8	Q9tvg8 plasmodium
45	845	41.0	1726	5	Q02569	Q02569 plasmodium

ALIGNMENTS

RESULT 1
Q25924
ID Q25924 PRELIMINARY; PRT; 651 AA.
AC Q25924;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE MEROZOITE SURFACE ANTIGEN 1 (FRAGMENT).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RO-71;
RX MEDLINE=92275047; PubMed=1592091;
RA Olafsson P., Matile H., Certa U.;
RT "Plasmodium falciparum: the repetitive MSA-1 surface protein of the
RT RO-71 isolate is recognized by mouse antibody against the
RT nonrepetitive repeat block of RO-33.";
RL Exp. Parasitol. 74:381-389(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RO-71;
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RT merozoite surface antigen-1.";
RL Exp. Parasitol. 81:47-54(1995).
DR EMBL; Z35329; CAA84558.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Merozoite.
FT NON_TER 1 1
SQ SEQUENCE 651 AA; 74134 MW; AA2137E699255150 CRC64;

Query Match 99.1%; Score 2043.5; DB 5; Length 651;
Best Local Similarity 99.5%; Pred. No. 2.2e-102;
Matches 392; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Qy 1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db    259 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 318

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 378

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    379 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 437

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    438 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 497

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db    498 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPQNSGCFR 557

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    558 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 617

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||
Db    618 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 651


RESULT  2
Q25983
ID   Q25983         PRELIMINARY;   PRT;   569 AA.
AC   Q25983;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13355; BAA02616.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER      1      1
SQ   SEQUENCE   569 AA;  64491 MW;  57A6B62FF72CE885 CRC64;


  Query Match             98.8%;  Score 2036.5;  DB 5;  Length 569;
  Best Local Similarity   99.5%;  Pred. No. 4.6e-102;
  Matches  392;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 236

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 535

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||
Db    536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT  3
Q25969
ID   Q25969         PRELIMINARY;   PRT;   569 AA.
AC   Q25969;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13346; BAA02607.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER      1      1
SQ   SEQUENCE   569 AA;  64485 MW;  A26C45914AF01396 CRC64;


  Query Match             98.5%;  Score 2030.5;  DB 5;  Length 569;
  Best Local Similarity   99.2%;  Pred. No. 9.7e-102;
  Matches  391;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 535

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||
Db    536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569
```

```
Db    536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT   4
Q25974
ID   Q25974         PRELIMINARY;   PRT;   569 AA.
AC   Q25974;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13348; BAA02609.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER      1      1
SQ   SEQUENCE   569 AA;  64457 MW;  A2691FE569E80396 CRC64;


  Query Match             98.5%;  Score 2030.5;  DB 5;  Length 569;
  Best Local Similarity   99.2%;  Pred. No. 9.7e-102;
  Matches  391;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 535

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db    536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT   5
Q25975
ID   Q25975         PRELIMINARY;   PRT;   569 AA.
AC   Q25975;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13349; BAA02610.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER      1      1
SQ   SEQUENCE   569 AA;  64459 MW;  ED4284B2867C9703 CRC64;


  Query Match             98.5%;  Score 2030.5;  DB 5;  Length 569;
  Best Local Similarity   99.2%;  Pred. No. 9.7e-102;
  Matches  391;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 535

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db    536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT   6
Q25977
ID   Q25977         PRELIMINARY;   PRT;   569 AA.
AC   Q25977;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13350; BAA02611.1; -.
```

```
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64485 MW;  A26C45914AF01396 CRC64;


  Query Match             98.5%;  Score 2030.5;  DB 5;  Length 569;
  Best Local Similarity   99.2%;  Pred. No. 9.7e-102;
  Matches  391;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 535

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db    536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT   7
Q25979
ID   Q25979         PRELIMINARY;   PRT;   569 AA.
AC   Q25979;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13352; BAA02613.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64457 MW;  A2691FE569E80396 CRC64;


  Query Match             98.5%;  Score 2030.5;  DB 5;  Length 569;
  Best Local Similarity   99.2%;  Pred. No. 9.7e-102;
  Matches  391;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 535

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db    536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569


RESULT   8
Q9TZT5
ID   Q9TZT5      PRELIMINARY;      PRT;  1694 AA.
AC   Q9TZT5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR.
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HN1;
RA   Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF062348; AAC72884.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite.
SQ   SEQUENCE   1694 AA;  192794 MW;  84CFC0E709F5673B CRC64;


  Query Match             98.5%;  Score 2030.5;  DB 5;  Length 1694;
  Best Local Similarity   99.2%;  Pred. No. 3e-101;
  Matches  391;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1302 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 1361

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1362 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1421

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   1422 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1480

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   1481 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1540

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1541 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1600

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1601 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1660

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db   1661 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1694


RESULT  9
Q9NHX1
ID   Q9NHX1      PRELIMINARY;      PRT;  1694 AA.
AC   Q9NHX1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MAJOR MEROZOITE SURFACE ANTIGEN.
GN   GP195.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FCC1/HN;
RA   Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;
RT   "Molecular cloning and sequence analysis of major merozoite surface
RT   antigen(gp195)gene of Plasmodium falciparum isolate FCC1/HN.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF218248; AAF27526.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR003247; -.
DR   Pfam; PF00008; EGF; 1.
DR   ProDom; PD001527; -; 1.
KW   Merozoite.
SQ   SEQUENCE   1694 AA;  192766 MW;  B51634A49E0F6728 CRC64;


  Query Match             98.5%;  Score 2030.5;  DB 5;  Length 1694;
  Best Local Similarity   99.2%;  Pred. No. 3e-101;
  Matches  391;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1302 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 1361

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1362 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1421

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   1422 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1480

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   1481 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1540

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1541 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1600

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1601 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1660

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db   1661 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1694


RESULT  10
Q9TZT4
ID   Q9TZT4      PRELIMINARY;      PRT;  1704 AA.
AC   Q9TZT4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR.
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HN2;
RA   Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF062349; AAC72885.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite.
SQ   SEQUENCE   1704 AA;  193762 MW;  385526D0DA56FD1D CRC64;


  Query Match             98.5%;  Score 2030.5;  DB 5;  Length 1704;
  Best Local Similarity   99.2%;  Pred. No. 3.1e-101;
  Matches  391;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1312 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 1371

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1372 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1431

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   1432 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1490

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   1491 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1550

Qy    241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1551 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1610

Qy    301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1611 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1670

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||
Db   1671 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1704


RESULT  11
Q25968
ID   Q25968         PRELIMINARY;   PRT;   570 AA.
AC   Q25968;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
```

```
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13345; BAA02606.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   570 AA;  64632 MW;  424BF553CCC2F2BE CRC64;


  Query Match             98.3%;  Score 2027;  DB 5;  Length 570;
  Best Local Similarity   98.7%;  Pred. No. 1.5e-101;
  Matches  389;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db     177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     297 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 356

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     357 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 416

Qy     241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     417 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 476

Qy     301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
           |||||||||||||||||||||||||||||||||||||||| |||||||: |||||||||
Db     477 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 536

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           ||||||||||||||||||||||||||||||||||
Db     537 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 570


RESULT  12
Q9TYG2
ID   Q9TYG2      PRELIMINARY;      PRT;   570 AA.
AC   Q9TYG2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88011243; PubMed=3079521;
RA   Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT   "Allelic dimorphism in a surface antigen gene of the malaria parasite
RT   Plasmodium falciparum.";
RL   J. Mol. Biol. 195:273-287(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13343; BAA02604.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   570 AA;  64630 MW;  8674DEC89B2D662A CRC64;


  Query Match             98.3%;  Score 2027;  DB 5;  Length 570;
  Best Local Similarity   98.7%;  Pred. No. 1.5e-101;
  Matches  389;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db     177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 236

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     297 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 356

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     357 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 416

Qy     241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     417 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 476

Qy     301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
           |||||||||||||||||||||||||||||||||||||||| |||||||: |||||||||
Db     477 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 536

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           ||||||||||||||||||||||||||||||||||
Db     537 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 570


RESULT  13
Q25922
ID   Q25922      PRELIMINARY;      PRT;  1720 AA.
AC   Q25922;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PRECURSOR OF THE MAJOR MEROZOITE SURFACE ANTIGENS.
OS   Plasmodium falciparum (isolate NF54).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5843;
RN   [1]
RP   SEQUENCE OF 1069-1720 FROM N.A.
RC   STRAIN=NF54;
RA   Tolle R., Bujard H., Cooper J.A.;
RL   Exp. Parasitol. 0:0-0(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NF54;
RA   Tolle R.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NF54;
RX   MEDLINE=96123395; PubMed=8577332;
RA   Pan W., Tolle R., Bujard H.;
RT   "A direct and rapid sequencing strategy for the Plasmodium falciparum
RT   antigen gene gp190/MSA1.";
RL   Mol. Biochem. Parasitol. 73:241-244(1995).
DR   EMBL; Z35327; CAA84556.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Signal; Merozoite.
```

```
FT   SIGNAL        1     19       POTENTIAL.
SQ   SEQUENCE   1720 AA;  195725 MW;  717B2FD1E637A8A3 CRC64;


  Query Match             98.2%;  Score 2025;  DB 5;  Length 1720;
  Best Local Similarity   98.7%;  Pred. No. 6.1e-101;
  Matches  389;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1327 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 1386

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1387 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1446

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1447 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1506

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||  |||||||
Db    1507 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFEAIKK 1566

Qy     241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1567 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1626

Qy     301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
           |||||||||||||||||||||||||||||||||||| |||||||: |||||||||
Db    1627 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 1686

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db    1687 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1720


RESULT  14
Q25923
ID   Q25923      PRELIMINARY;      PRT;   652 AA.
AC   Q25923;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MEROZOITE SURFACE ANTIGEN 1 (FRAGMENT).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE OF 1-298 FROM N.A.
RC   STRAIN=FCH5/NF7;
RX   MEDLINE=88142999; PubMed=2449612;
RA   Peterson G.M., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
RA   Brown G.V., Anders R.F., Kemp D.J.;
RT   "Variation in the precursor to the major merozoite surface antigens of
RT   Plasmodium falciparum.";
RL   Mol. Biochem. Parasitol. 27:291-302(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FCH5/NF7;
RA   Tolle R.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FCH5/NF7;
RX   MEDLINE=95354793; PubMed=7628566;
RA   Tolle R., Bujard H., Cooper J.A.;
RT   "Plasmodium falciparum: variations within the C-terminal region of
RT   merozoite surface antigen-1.";
RL   Exp. Parasitol. 81:47-54(1995).
DR   EMBL; Z35328; CAA84557.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite.
FT   NON_TER       1      1
SQ   SEQUENCE   652 AA;  74292 MW;  2B6A87737B490A62 CRC64;


  Query Match             98.0%;  Score 2021;  DB 5;  Length 652;
  Best Local Similarity   98.5%;  Pred. No. 3.6e-101;
  Matches  388;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 318

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 378

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     379 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 438

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||  |||||||
Db     439 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFEAIKK 498

Qy     241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 558

Qy     301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
           |||||||||||||||||||||||||||||||||||| |||:|||: |||||||||
Db     559 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEKDSGSSRKKITCECTK 618

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db     619 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 652


RESULT  15
Q25978
ID   Q25978         PRELIMINARY;   PRT;   569 AA.
AC   Q25978;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN   MSP1.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93295445; PubMed=8515786;
RA   Jongwutiwes S., Tanabe K., Kanbara H.;
RT   "Sequence conservation in the C-terminal part of the precursor to the
RT   major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT   field isolates.";
RL   Mol. Biochem. Parasitol. 59:95-100(1993).
DR   EMBL; D13351; BAA02612.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Merozoite; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   569 AA;  64536 MW;  B8B9B63EECB1DA51 CRC64;


  Query Match             97.8%;  Score 2017.5;  DB 5;  Length 569;
  Best Local Similarity   98.7%;  Pred. No. 4.8e-101;
  Matches  389;  Conservative    1;  Mismatches    3;  Indels    1;  Gaps    1;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 236
```

```
Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     237 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 296

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db     297 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 355

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     356 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

Qy     241 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     416 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 475

Qy     301 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
           |||||||||||||||||||||||||||||||||||| |||||||: |||||||||
Db     476 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 535

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db     536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569
```

Search completed: August 8, 2001, 12:33:42
Job time: 270 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:32:29 ; Search time 24.79 Seconds
(without alignments)
544.440 Million cell updates/sec

Title: US-09-500-376-3
Perfect score: 2052
Sequence: 1 AISVTMDNILSGFENEYDVI.....SNFLGISFLLILMLILYSFI 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2033	99.1	1701	1	MSP1_PLAFF	P13819 plasmodium
2	2033	99.1	1701	1	MSP1_PLAFM	P08569 plasmodium
3	2006	97.8	1726	1	MSP1_PLAFC	P04934 plasmodium
4	2003	97.6	1726	1	MSP1_PLAFP	P50495 plasmodium
5	1980.5	96.5	1682	1	MSP1_PLAF3	P19598 plasmodium
6	1254.5	61.1	1630	1	MSP1_PLAFK	P04932 plasmodium
7	1254.5	61.1	1639	1	MSP1_PLAFW	P04933 plasmodium
8	593.5	28.9	1772	1	MSP1_PLAYO	P13828 plasmodium
9	147	7.2	1130	1	YL17_CAEEL	Q11102 caenorhabdi
10	147	7.2	1679	1	YIO9_YEAST	P40457 saccharomyc
11	145.5	7.1	2116	1	MYS2_DICDI	P08799 dictyosteli
12	133	6.5	1228	1	ECM_HUMAN	Q13201 homo sapien
13	132.5	6.5	1875	1	MLP1_YEAST	Q02455 saccharomyc
14	132	6.4	944	1	NUF1_YEAST	P32380 saccharomyc
15	130.5	6.4	1038	1	CIN8_YEAST	P27895 saccharomyc
16	130	6.3	1170	1	SMC2_YEAST	P38989 saccharomyc
17	130	6.3	2869	1	RBP1_PLAVB	Q00798 plasmodium
18	129	6.3	1790	1	USO1_YEAST	P25386 saccharomyc
19	128.5	6.3	1312	1	RA50_YEAST	P12753 saccharomyc
20	127.5	6.2	2710	1	TOXA_CLODI	P16154 clostridium
21	127	6.2	540	1	YKZ6_YEAST	P36112 saccharomyc
22	127	6.2	996	1	PPOL_SARPE	Q11208 sarcophaga
23	126.5	6.2	798	1	ITB1_HUMAN	P05556 homo sapien
24	126	6.1	796	1	YFC3_YEAST	P43573 saccharomyc
25	125.5	6.1	1526	1	MYS2_SCHPO	Q9usi6 schizosacch
26	125.5	6.1	1805	1	HMW2_MYCGE	P47460 mycoplasma
27	124.5	6.1	743	1	ABRA_PLAFC	P22620 plasmodium
28	124	6.0	729	1	KAR3_YEAST	P17119 saccharomyc
29	123.5	6.0	622	1	YAE7_YEAST	P39723 saccharomyc
30	123.5	6.0	984	1	NOF_DROME	P16320 drosophila
31	122	5.9	600	1	ABRA_PLAFG	P23745 plasmodium
32	122	5.9	1251	1	RBP2_PLAVB	Q00799 plasmodium
33	121.5	5.9	707	1	YJ9C_YEAST	P47166 saccharomyc
34	121.5	5.9	1928	1	MYS1_YEAST	P08964 saccharomyc
35	121.5	5.9	2748	1	NUM1_YEAST	Q00402 saccharomyc
36	121	5.9	512	1	YE94_SCHPO	O13766 schizosacch
37	121	5.9	1957	1	YD86_SCHPO	Q10411 schizosacch
38	120.5	5.9	506	1	CC37_YEAST	P06101 saccharomyc
39	120	5.8	399	1	RMS5_PENUR	P47907 penicillium
40	120	5.8	594	1	UVRC_HELPJ	Q9z121 helicobacte
41	120	5.8	1230	1	SMC3_YEAST	P47037 saccharomyc
42	120	5.8	2245	1	MYSJ_DICDI	P54697 dictyosteli
43	119	5.8	808	1	Y066_NPVAC	P41467 autographa
44	119	5.8	1030	1	Y018_MYCPN	P75093 mycoplasma
45	119	5.8	1169	1	EX5B_BORBU	O51578 borrelia bu

ALIGNMENTS

```
RESULT   1
MSP1_PLAFF
ID   MSP1_PLAFF     STANDARD;      PRT;  1701 AA.
AC   P13819;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA).
GN   MSP-1.
OS   Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88142999; PubMed=2449612;
RA   Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
RA   Brown G.V., Anders R.F., Kemp D.J.;
RT   "Variation in the precursor to the major merozoite surface antigens
RT   of Plasmodium falciparum.";
RL   Mol. Biochem. Parasitol. 27:291-302(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19143; AAA29653.1; -.
DR   PIR; A54498; A54498.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1701       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    536    536       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    607    607       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    802    802       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    899    899       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    965    965       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    991    991       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1089   1089       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1196   1196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1588   1588       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
SQ   SEQUENCE   1701 AA;  193719 MW;  3920B75E73D38552 CRC64;


  Query Match             99.1%;  Score 2033;  DB 1;  Length 1701;
  Best Local Similarity   99.2%;  Pred. No. 2.7e-97;
  Matches  391;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1308 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 1367

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    1368 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1427

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1428 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1487

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1488 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1547

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1548 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1607

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1608 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 1667

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db    1668 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1701


RESULT   2
MSP1_PLAFM
ID   MSP1_PLAFM     STANDARD;      PRT;  1701 AA.
AC   P08569;
DT   01-AUG-1988 (Rel. 08, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P190).
GN   MSP-1.
OS   Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=70153;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88011243; PubMed=3079521;
RA   Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT   "Allelic dimorphism in a surface antigen gene of the malaria parasite
RT   Plasmodium falciparum.";
RL   J. Mol. Biol. 195:273-287(1987).
RN   [2]
RP   REVISIONS TO 1403; 1569 AND 1629.
RA   Tanabe K.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-115 FROM N.A.
RX   MEDLINE=86136024; PubMed=3004972;
RA   Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA   Stunnenberg H., Bujard H.;
RT   "Polymorphism of the precursor for the major surface antigens of
RT   Plasmodium falciparum merozoites: studies at the genetic level.";
RL   EMBO J. 4:3823-3829(1985).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X05624; CAA29112.1; -.
DR   PIR; A26868; A26868.
DR   PIR; B25120; B25120.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1701       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    536    536       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    607    607       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    802    802       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    899    899       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    965    965       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    991    991       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1089   1089       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1196   1196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1588   1588       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1701 AA;  193768 MW;  3FC2EC59AF96EA98 CRC64;


  Query Match             99.1%;  Score 2033;  DB 1;  Length 1701;
  Best Local Similarity   99.2%;  Pred. No. 2.7e-97;
  Matches  391;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1308 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 1367

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    1368 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1427

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1428 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1487

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1488 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1547

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1548 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1607

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1608 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 1667

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||||
Db    1668 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1701


RESULT   3
MSP1_PLAFC
ID   MSP1_PLAFC     STANDARD;      PRT;  1726 AA.
AC   P04934;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
```

```
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P195).
GN   MSP-1.
OS   Plasmodium falciparum (isolate Camp / Malaysia).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5835;
RN   [1]
RP   SEQUENCE OF 1-1103 FROM N.A.
RX   MEDLINE=86205236; PubMed=3517809;
RA   Weber J.L., Leininger W.M., Lyon J.A.;
RT   "Variation in the gene encoding a major merozoite surface antigen of
RT   the human malaria parasite Plasmodium falciparum.";
RL   Nucleic Acids Res. 14:3311-3323(1986).
RN   [2]
RP   SEQUENCE OF 1104-1726 FROM N.A.
RX   MEDLINE=88143999; PubMed=3278296;
RA   Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
RT   "Merozoite surface protein sequence from the Camp strain of the human
RT   malaria parasite Plasmodium falciparum.";
RL   Nucleic Acids Res. 16:1206-1206(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03831; CAA27446.1; -.
DR   PIR; A23386; SAZQGM.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1726       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    133    133       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    567    567       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    638    638       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    924    924       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    944    944       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    990    990       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1016   1016       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1114   1114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1221   1221       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1613   1613       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1726 AA;  196197 MW;  DD8AD45FA352BCF3 CRC64;


  Query Match             97.8%;  Score 2006;  DB 1;  Length 1726;
  Best Local Similarity   98.0%;  Pred. No. 6.6e-96;
  Matches  386;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    1333 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 1392

Qy      61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    1393 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1452

Qy     121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    1453 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1512

Qy     181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    1513 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 1572

Qy     241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    1573 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1632

Qy     301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSSRKKITCECTK 360
           |||||||||||||||||||| |||||||||||||||| |||||||: |||||||||
Db    1633 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1692

Qy     361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           |||||||||||||||||||||||||||||||||
Db    1693 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   4
MSP1_PLAFP
ID   MSP1_PLAFP     STANDARD;      PRT;  1726 AA.
AC   P50495;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (GP195).
GN   MSP-1.
OS   Plasmodium falciparum (isolate Palo Alto / Uganda).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=57270;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89005525; PubMed=3049134;
RA   Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
RA   Siddiqui W.A.;
RT   "Plasmodium falciparum: gene structure and hydropathy profile of the
RT   major merozoite surface antigen (gp195) of the Uganda-Palo Alto
RT   isolate.";
RL   Exp. Parasitol. 67:1-11(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M37213; AAA29611.1; -.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1726       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    133    133       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    501    501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    567    567       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    638    638       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    924    924       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    944    944       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    990    990       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1016   1016       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1114   1114       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD   1221   1221       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1613   1613       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1726 AA;  196174 MW;  5B59CEEFA2F9A026 CRC64;


  Query Match             97.6%;  Score 2003;  DB 1;  Length 1726;
  Best Local Similarity   97.7%;  Pred. No. 9.4e-96;
  Matches  385;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1333 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 1392

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1393 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1452

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1453 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1512

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||:||||
Db   1513 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFDAIKK 1572

Qy    241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1573 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 1632

Qy    301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
          |||||||||||||||||| |||||||||||||||| |||||||: |||||||||
Db   1633 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1692

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||||
Db   1693 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1726


RESULT   5
MSP1_PLAF3
ID   MSP1_PLAF3     STANDARD;      PRT;  1682 AA.
AC   P19598; Q25921;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P190).
GN   MSP-1.
OS   Plasmodium falciparum (isolate ro-33 / Ghana).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5834;
RN   [1]
RP   SEQUENCE OF 1-1061 FROM N.A.
RX   MEDLINE=88166657; PubMed=3327688;
RA   Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT   "A naturally occurring gene encoding the major surface antigen
RT   precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
RL   EMBO J. 6:4137-4142(1987).
RN   [2]
RP   SEQUENCE OF 1032-1682 FROM N.A.
RX   MEDLINE=95354793; PubMed=7628566;
RA   Tolle R., Bujard H., Cooper J.A.;
RT   "Plasmodium falciparum: variations within the C-terminal region of
RT   merozoite surface antigen-1.";
RL   Exp. Parasitol. 81:47-54(1995).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M35727; AAA29715.1; -.
DR   EMBL; Y00087; CAA68280.1; -.
DR   EMBL; Z35326; CAA84555.1; -.
DR   PIR; S06286; S06286.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1682       MEROZOITE SURFACE PROTEIN 1.
FT   TRANSMEM   1666   1682       MEMBRANE ANCHOR (POTENTIAL).
FT   CARBOHYD    233    233       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    462    462       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    528    528       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    599    599       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    785    785       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    881    881       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    901    901       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    947    947       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1071   1071       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1178   1178       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1569   1569       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1682 AA;  192462 MW;  C82A1E159948CAD6 CRC64;


  Query Match             96.5%;  Score 1980.5;  DB 1;  Length 1682;
  Best Local Similarity   97.2%;  Pred. No. 1.3e-94;
  Matches  383;  Conservative    2;  Mismatches    8;  Indels    1;  Gaps    1;

Qy      1 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKY 60
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1290 AISVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKY 1349

Qy     61 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGIS 120
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1350 FLDVLESDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1409

Qy    121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 180
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   1410 YYEKVLAKYKDDLESIKKVIKEEKE-FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETL 1468

Qy    181 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||  |||||||
Db   1469 YNNLVNKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFEAIKK 1528

Qy    241 LINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR 300
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1529 LINDDTKKDMLGKLLSTGLVQNFPNTIISKLIEGKFQDMLNISQHQCVKKQCPQNSGCFR 1588

Qy    301 HLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
          |||||||||||||||||| |||||||||||||||| |||||||: |||||||||
Db   1589 HLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1648

Qy    361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
          |||||||||||||||||||||||||||||||||||
Db   1649 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1682


RESULT   6
MSP1_PLAFK
ID   MSP1_PLAFK     STANDARD;      PRT;  1630 AA.
AC   P04932;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P190).
```

```
GN   MSP-1.
OS   Plasmodium falciparum (isolate K1 / Thailand).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5839;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86136024; PubMed=3004972;
RA   Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA   Stunnenberg H., Bujard H.;
RT   "Polymorphism of the precursor for the major surface antigens of
RT   Plasmodium falciparum merozoites: studies at the genetic level.";
RL   EMBO J. 4:3823-3829(1985).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RA   Pan W., Tolle R., Bujard H.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03371; CAA27070.1; -.
DR   PIR; A25120; SAZQK1.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1630       MEROZOITE SURFACE PROTEIN 1.
FT   DOMAIN       67     84       TRIPEPTIDE SG(TP) REPEAT.
FT   TRANSMEM   1614   1630       MEMBRANE ANCHOR.
FT   CARBOHYD     97     97       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    755    755       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    759    759       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    774    774       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    835    835       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    911    911       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    955    955       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1049   1049       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1156   1156       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1165   1165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1436   1436       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1517   1517       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1630 AA;  187289 MW;  ADBDEC3CE0A46322 CRC64;


  Query Match           61.1%;  Score 1254.5;  DB 1;  Length 1630;
  Best Local Similarity  61.2%;  Pred. No. 2.1e-57;
  Matches  240;  Conservative   54;  Mismatches   73;  Indels   25;  Gaps    3;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  ||||:|:|||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db    1261 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 1320

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |::: | || |||:|:: || || : : ||: :
Db    1321 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 1380

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                    |:| :  |:|||| ||||||  :
Db    1381 SEKYKSDLDSIKKYI--------------------NDKQGENEKYLPFLNNIETLYKTVN 1420

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
           :||| ::|:|:|:|: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db    1421 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 1480

Qy     246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHL 302
             ::| | ||||:|  | |   |::| |::|  | |||||||||||||:||||||||
Db    1481 NHNNLLTKFLSTGMV--FENLAKTVLSNLLDGNLQGMLNISQHQCVKKQCPQNSGCFRHL 1538

Qy     303 DEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPD 362
           |||||||||||||||||| |||||||||||||||| |||||||: ||||||||||
Db    1539 DEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPD 1598

Qy     363 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           ||||||||||||||||||||||||||||||||
Db    1599 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1630


RESULT   7
MSP1_PLAFW
ID   MSP1_PLAFW     STANDARD;      PRT;  1639 AA.
AC   P04933;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P195).
GN   MSP-1.
OS   Plasmodium falciparum (isolate Wellcome).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5848;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86014355; PubMed=2995820;
RA   Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA   Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA   Freeman R.R.;
RT   "Primary structure of the precursor to the three major surface
RT   antigens of Plasmodium falciparum merozoites.";
RL   Nature 317:270-273(1985).
RN   [2]
RP   REVISIONS.
RA   Holder A.A.;
RL   Submitted (MAR-1991) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X02919; CAA26676.1; -.
DR   PIR; A24594; A24594.
DR   InterPro; IPR000561; -.
DR   Pfam; PF00008; EGF; 1.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1639       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    116    116       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    268    268       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    764    764       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    768    768       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    783    783       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    844    844       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    920    920       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    964    964       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1058   1058       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD   1165   1165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1174   1174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1445   1445       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1526   1526       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1639 AA;  187618 MW;  2C255B6616C87F6E CRC64;


  Query Match             61.1%;  Score 1254.5;  DB 1;  Length 1639;
  Best Local Similarity   61.2%;  Pred. No. 2.1e-57;
  Matches  240;  Conservative   54;  Mismatches   73;  Indels   25;  Gaps    3;

Qy       6 MDNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVL 65
           :|||||  |||||:|:|||||||||||||||||:| |::|||:|: ||||||  ||: | :||
Db    1270 IDNILSKIENEYEVLYLKPLAGVYRSLKKQLENNVMTFNVNVKDILNSRFNKRENFKNVL 1329

Qy      66 ESDLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKV 125
           ||||: :| ::|: |:::| :| || |:::  | || |||:|:: || || : : ||: :
Db    1330 ESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFANDVLGYYKIL 1389

Qy     126 LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
             ||| ||:|||| |                |:| :  |:|||| |||||  :
Db    1390 SEKYKSDLDSIKKYI-------------------NDKQGENEKYLPFLNNIETLYKTVN 1429

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
           :||| ::|:|:||: :   ||    ||| :|: || | ||:  ||  |:|  |  |  |
Db    1430 DKIDLFVIHLEAKVLNYTYEKSNVEVKIKELNYLKTIQDKLADFKKNNNFVGIADLSTDY 1489

Qy     246 TKKDMLGKLLSTGLVQIFPN---TIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHL 302
             ::| | ||||:|  | |   |::| |::|  | ||||||||||||||:||||||||
Db    1490 NHNNLLTKFLSTGMV--FENLAKTVLSNLLDGNLQGMLNISQHQCVKKQCPQNSGCFRHL 1547

Qy     303 DEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTKPD 362
           |||||||||||||||||| |||||||||||||||| |||||||: ||||||||||
Db    1548 DEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPD 1607

Qy     363 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
           ||||||||||||||||||||||||||||||||
Db    1608 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1639


RESULT   8
MSP1_PLAYO
ID   MSP1_PLAYO     STANDARD;      PRT;  1772 AA.
AC   P13828;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (230 KDA).
GN   MSP-1.
OS   Plasmodium berghei yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5862;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90205979; PubMed=2320061;
RA   Lewis A.P.;
RT   "Sequence analysis upstream of the gene encoding the precursor to the
RT   major merozoite surface antigens of Plasmodium yoelii.";
RL   Mol. Biochem. Parasitol. 39:285-288(1990).
RN   [2]
RP   SEQUENCE OF 1093-1772 FROM N.A.
RC   STRAIN=17XL;
RX   MEDLINE=88124889; PubMed=2448778;
RA   Burns J.M. Jr., Daly T.M., Vaidya A.B., Long C.A.;
RT   "The 3' portion of the gene for a Plasmodium yoelii merozoite surface
RT   antigen encodes the epitope recognized by a protective monoclonal
RT   antibody.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:602-606(1988).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC       KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC       MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03612; AAA29762.1; -.
DR   EMBL; J04668; AAA29702.1; -.
DR   PIR; A28121; A28121.
DR   PIR; A45532; A45532.
KW   Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW   Transmembrane; GPI-anchor.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   1772       MEROZOITE SURFACE PROTEIN 1.
FT   CARBOHYD    54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   406    406       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   646    646       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   829    829       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1018   1018       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1090   1090       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1408   1408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1446   1446       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1541   1541       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1629   1629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1680   1680       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT  1521   1521       L -> V (IN REF. 2).
SQ   SEQUENCE   1772 AA;  197230 MW;  9A6291658EB0F45D CRC64;


  Query Match             28.9%;  Score 593.5;  DB 1;  Length 1772;
  Best Local Similarity   32.5%;  Pred. No. 1.8e-23;
  Matches  128;  Conservative   96;  Mismatches  137;  Indels   33;  Gaps    8;

Qy       8 NILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLES 67
           :||| | ||   :| | |   |:|||| : :   |   :: : ||:: :|| ||:||
Db    1401 DILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQKRNDFLEVLSH 1460

Qy      68 DLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKVLA 127
           :|  || :|:|:|:| : ::||::::|:  : : ||  : :  ||:   :|| :: |::
Db    1461 ELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFNKMVE 1520

Qy     128 KYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDEQKKE--SKFLPFLTNIETLYNNLV 185
             |   | ::|: |   :              |:|::  ||   |::| | ::: ||  ::
Db    1521 LYNTQLAAVKEQIATIE-------------AETNDTNKEEKKKYIPILEDLKGLYETVI 1566

Qy     186 NKIDDYLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLF----KNTNDFEAIKKL 241
           : ::|   |: ::::   || |  :    |     ||:|:| |    :|     :|
Db    1567 GQAEEYSEELQNRLDNYKNEKAEFEILTKNLEKYIQIDEKLDEFVEHAENNKHIASI--A 1624

Qy     242 INDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNIS-QHQCV-KKQCPENSGCF 299
           :|:  |  ::|:  |          |::|::     |:| :  :| ||  :  |:|:|||
Db    1625 LNNLNKSGLVGEGES--------KKILAKMLNMDGMDLLGVDPKHVCVDTRDIPKNAGCF 1676

Qy     300 RHLDEREECKCLLNYKQ-EGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCEC 358
           |  :  || :||| ||: ||: | || ||||: ||||||  |:|   :|  : ||| | |
Db    1677 RDDNGTEEWRCLLGYKKGEGNTCVENNNPTCDINNGGCDPTASCQNAESTENSKKIICTC 1736

Qy     359 TKPDSYPLFDGIFCSSSNFLGISFLLILMLILYS 392
           :|     ::|:|||||:|:|:| |||: ||:::
Db    1737 KEPTPNAYYEGVFCSSSSFMGLSILLIITLIVFN 1770


RESULT   9
YL17_CAEEL
ID   YL17_CAEEL     STANDARD;      PRT;  1130 AA.
AC   Q11102;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
```

```
DE   HYPOTHETICAL 131.5 KDA PROTEIN C02F12.7 IN CHROMOSOME X.
GN   C02F12.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Miller N.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: WEAK, TO MYOSINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U41545; AAA83190.1; -.
DR   WormPep; C02F12.7; CE03901.
KW   Hypothetical protein; Coiled coil.
FT   DOMAIN      121    779       COILED COIL (POTENTIAL).
FT   DOMAIN      805   1061       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1130 AA;  131485 MW;  B0FD2EFE3D99FB09 CRC64;


  Query Match             7.2%;  Score 147;  DB 1;  Length 1130;
  Best Local Similarity   23.6%;  Pred. No. 0.9;
  Matches   77;  Conservative   64;  Mismatches  129;  Indels   56;  Gaps   14;

Qy      31 SLKKQIEKNIITFNLNLNDILNSRLKK--------RKYFLDVLESDLMQFKHISSN--EY 80
           :|:||:||  |:   : | :|:   :|           : |||::: |||    |  ||
Db     596 NLRKQLEKE-ISHTEDRNRLLHENTQKELEAHKETHTETVRVLEAEIDQFKSAFENEQEY 654

Qy      81 IIEDSFKLLNSEQKN-ILLKSYKYIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKV 139
             | | |:   | :|  ||   : :|   ||   |  :  :  |:: :| | ::| :|:
Db     655 GKEKSAKIRELEAQNKTLLSEMEKVKHVAENLEAFTSDKDNLLEELESKNK-NIEHLKQE 713

Qy     140 IKEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKIDDYLINLKAKI 199
           | :  ||              | | :|:|:    :  :|   ::  ::|:   : :   :
Db     714 IAQLNEKI------------STKETEKDSELEKTIAQLEIDNSSKSDQIEKLHLRVNDML 761

Qy     200 NDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKL--INDDTKKDMLGKLLST 257
           :     |||    : |  ::| |  |      :|  |:  ||  : :: :|:
Db     762 DQMGTIKDEL---VKKNEEIKTISAKTAQLLESNTVESETKLASVTEEREKE-------- 810

Qy     258 GLVQIFPNTIISKLIEGKFQDMLNISQ-----HQCVKKQCPENSGCFRHLDE-----REE 307
             :| |  | ||:    |  ::|  ::       |:|::  | :|    ||:      ||
Db     811 --IQSF-QTQISE----KDNEVLTKAERINELETCLKEREVELTGMRTKLDDMTQQLNEE 863

Qy     308 CKCLLNYKQEGDKCEENPNPTCNENN 333
             :|      :| :|    | || |
Db     864 TTVVLFDNSIQEKIDEK-EATINEMN 888


RESULT  10
YIO9_YEAST
ID   YIO9_YEAST     STANDARD;      PRT;  1679 AA.
AC   P40457;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   HYPOTHETICAL 195.1 KDA PROTEIN IN DNA43-UBI1 INTERGENIC REGION.
GN   YIL149C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z38059; CAA86129.1; -.
DR   PIR; S48385; S48385.
DR   SGD; S0001411; MLP2.
KW   Hypothetical protein.
SQ   SEQUENCE   1679 AA;  195141 MW;  298950CC52202D8F CRC64;


  Query Match             7.2%;  Score 147;  DB 1;  Length 1679;
  Best Local Similarity   22.7%;  Pred. No. 1.4;
  Matches  105;  Conservative   70;  Mismatches  150;  Indels  138;  Gaps   22;

Qy       3 SVTMDNILSGFENEYDVIYLKPLAGVYRSLKKQ---IEKNIITFNLNLNDILNSRLKKRK 59
           || :: :    |  |  :|| |  :: :|: :   :|  : |    |     : :|::
Db    1079 SVLIEKVDDTAANNGDKDHLK-LVSLFSNLRHERNSLETKLTTCKREL-----AFVKQKN 1132

Qy      60 YFLDVLESDLMQFKHISSNEY----IIEDSFKLLNSE--QKNILLKSYKYIKESVEN--- 110
             |:   :|| : : :|  ||    :| | || :  |  | ||| ::   :::|::|
Db    1133 DSLEKTINDLQRTQTLSEKEYQCSAVIIDEFKDITKEVTQVNILKENNAILQKSLKNVTE 1192

Qy     111 -------DIKFAQEGISYYEKVLAKYKDDLE-SIKKVIKEEKEKFPSSPPTTPPSPAKTD 162
                :   || ||  :: | : |: :  :  |::  | |           |  : |
Db    1193 KNREIYKQLNDRQEEISRLQRDLIQTKEQVSINSNKILVYESEMEQCKQRYQDLSQQQKD 1252

Qy     163 EQKKESKFLPFLTNIETLYNNLVNKIDDYLINLKAKI-----------NDCNVEKDEAHV 211
           |||:                | |:| |    || |:           |  |  | :||
Db    1253 AQKKD-------------IEKLTNEISD----LKGKLSSAENANADLENKFNRLKKQAHE 1295

Qy     212 KI-----------TKLSDLKAIDDKI--DL-FKNTNDFEAIKKL---------INDDTKK 248
           |:              :|::|||| ||:  || |:|    :   ||         :: | :|
Db    1296 KLDASKKQQAALTNELNELKAIKDKLEQDLHFENAKVIDLDTKLKAHELQSEDVSRDHEK 1355

Qy     249 DMLGKLLS-----TGLVQIF-------------------PNTII---SKLIEGKFQDML 280
           |    |:          :|||                    : ||   :|  | | |: |
Db    1356 DTYRTLMEEIESLKKELQIFKTANSSSDAFEKLKVNMEKEKDRIIDERTKEFEKKLQETL 1415

Qy     281 NISQH---------QCVKKQCPE--NSGCFRHLDEREE-------------CKCLLNYK 315
           | |            : :||:  :       | : | ||               | :   |
Db    1416 NKSTSSEAEYSKDIETLKKEWLKEYEDETLRRIKEAEENLKKRIRLPSEERIQKIISKRK 1475

Qy     316 QEGD-----KCEENPNPTCNENNGGCDADATCTEEDSGSSRKK 353
           :| :     | :||       :| |   ||   ||:  :|  |
Db    1476 EELEEEFRKKLKENAGSLTFLDNKGSGEDA---EEELWNSPSK 1515


RESULT  11
MYS2_DICDI
ID   MYS2_DICDI     STANDARD;      PRT;  2116 AA.
AC   P08799;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   MYOSIN II HEAVY CHAIN, NON MUSCLE.
GN   MHCA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87092266; PubMed=3540939;
RA   Warrick H.M., de Lozanne A., Leinwand L.A., Spudich J.A.;
RT   "Conserved protein domains in a myosin heavy chain gene from
RT   Dictyostelium discoideum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN   [2]
RP   PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC   STRAIN=AX2;
RX   MEDLINE=90353583; PubMed=2387408;
RA   Lueck-Vielmeter D., Schleicher M., Grabatin B., Wippler J.,
RA   Gerisch G.;
RT   "Replacement of threonine residues by serine and alanine in a
RT   phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RL   FEBS Lett. 269:239-243(1990).
RN   [3]
RP   PHOSPHORYLATION SITES.
RX   MEDLINE=88112226; PubMed=2828113;
RA   Wagle G., Noegel A., Scheel J., Gerisch G.;
RT   "Phosphorylation of threonine residues on cloned fragments of the
RT   Dictyostelium myosin heavy chain.";
RL   FEBS Lett. 227:71-75(1988).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX   MEDLINE=95345066; PubMed=7619795;
RA   Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,
RA   Rayment I.;
RT   "X-ray structures of the myosin motor domain of Dictyostelium
RT   discoideum complexed with MgADP.BeFx and MgADP.AlF4-.";
RL   Biochemistry 34:8960-8972(1995).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX   MEDLINE=95345067; PubMed=7619796;
RA   Smith C.A., Rayment I.;
RT   "X-ray structure of the magnesium(II)-pyrophosphate complex of the
RT   truncated head of Dictyostelium discoideum myosin to 2.7-A
RT   resolution.";
RL   Biochemistry 34:8973-8981(1995).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX   MEDLINE=96206189; PubMed=8611530;
RA   Smith C.A., Rayment I.;
RT   "X-ray structure of the magnesium(II).ADP.vanadate complex of the
RT   Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";
RL   Biochemistry 35:5404-5417(1996).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX   MEDLINE=97452580; PubMed=9305951;
RA   Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT   "X-ray structures of the MgADP, MgATPgammaS, and MgAMPPNP complexes
RT   of the Dictyostelium discoideum myosin motor domain.";
RL   Biochemistry 36:11619-11628(1997).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX   MEDLINE=98070605; PubMed=9405148;
RA   Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
RT   "X-ray crystal structure and solution fluorescence characterization
RT   of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the
RT   Dictyostelium discoideum myosin motor domain.";
RL   J. Mol. Biol. 274:394-407(1997).
CC   -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
CC       ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC   -!- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
CC       INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
CC       LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
CC       (MLC-2).
CC   -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
CC       CORTEX.
CC   -!- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER
CC       SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
CC       THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC   -!- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
CC       ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
CC       POSITION (688).
CC   -!- SIMILARITY: THE GLOBULAR HEAD AA SEQUENCE SHOWS A HIGH DEGREE OF
CC       SIMILARITY WITH THE GLOBULAR HEAD SEQUENCES OF MUSCLE & NONMUSCLE
CC       HEAVY CHAINS. BY CONTRAST THE ROD SEQUENCE IS LESS CONSERVED, BUT
CC       THE PERIODICITIES OF HYDROPHOBIC & CHARGED RESIDUES, WHICH DICTATE
CC       THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14628; AAA33227.1; -.
DR   PIR; A26655; A26655.
DR   PIR; S00250; S00250.
DR   PDB; 1MMA; 03-DEC-97.
DR   PDB; 1MMD; 17-AUG-96.
DR   PDB; 1MMG; 03-DEC-97.
DR   PDB; 1MMN; 03-DEC-97.
DR   PDB; 1MND; 17-AUG-96.
DR   PDB; 1MNE; 17-AUG-96.
DR   PDB; 1VOM; 23-DEC-96.
DR   PDB; 1LVK; 28-JAN-98.
DR   DictyDb; DD01008; mhcA.
DR   InterPro; IPR000048; -.
DR   InterPro; IPR001609; -.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF00063; myosin_head; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
KW   Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
KW   Heptad repeat pattern; Methylation; Alkylation; Phosphorylation.
FT   DOMAIN        1    816       GLOBULAR HEAD (S1).
FT   DOMAIN      817   2116       RODLIKE TAIL (S2 AND LMM DOMAINS).
FT   DOMAIN      817   2116       COILED COIL (POTENTIAL).
FT   NP_BIND     179    186       ATP.
FT   DOMAIN      638    660       ACTIN-BINDING.
FT   DOMAIN      738    752       ACTIN-BINDING.
FT   MOD_RES     130    130       METHYLATION (DI-) (POTENTIAL).
FT   MOD_RES     678    678       ALKYLATION (SH-1).
FT   MOD_RES    1823   1823       PHOSPHORYLATION (BY MHCK).
FT   MOD_RES    1833   1833       PHOSPHORYLATION (BY MHCK).
FT   MOD_RES    2029   2029       PHOSPHORYLATION (BY MHCK).
SQ   SEQUENCE   2116 AA;  243871 MW;  2FC3770BB1EE56A1 CRC64;


  Query Match             7.1%;  Score 145.5;  DB 1;  Length 2116;
  Best Local Similarity   23.1%;  Pred. No. 2.1;
  Matches   82;  Conservative   61;  Mismatches  153;  Indels   59;  Gaps   12;

Qy      22 LKPLAGVYRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDV-LESDLMQFKHISSNEY 80
           :| |      :|:|:|  |:           :| | |  |:| ||    | :    :
Db    1673 IKSLVAEVDEVKEQLEDEILA---------KDKLVKAKRALEVELEEVRDQLEEEEDSRS 1723

Qy      81 IIEDSFKLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVI 140
             :||| : | :| ::|     ||  | || : |         ::   |  ||:::::|| :
Db    1724 ELEDSKRRLTTEVEDI---KKKYDAE-VEQNTKL--------DEAKKKLTDDVDTLKKQL 1771

Qy     141 KEEKEKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV------NKIDDYLIN 194
           ::||:|   |          :  ::: ||:   ||  ::    |          | :  | :
Db    1772 EDEKKKLNES---------ERAKKRLESENEDFLAKLDAEVKNRSRAEKDRKKYEKDLKD 1822

Qy     195 LKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKKDMLGKL 254
```

```
          | |:||    | :  :   ||      :|:||  ::  : |  |    | :|| : |:
Db   1823 TKYKLNDEAATKTQTEIGAAKL------EDQIDELRSKLEQEQAKATQADKSKKTLEGE- 1875

Qy    255 LSTGLVQIFPNTII-------SKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDEREE 307
          :    ||     |        : :||: :::    :     |   | |   : | | |
Db   1876 IDNLRAQIEDEGKIKMRLEKEKRALEGELEELRETVEEAEDSKSEAEQS---KRLVELEL 1932

Qy    308 CKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEE-----DSGSSRKKITCE 357
            | ::| |  |   :   |      :|     ||     :|  |||::  |
Db   1933 EDARRNLQKEIDAKEIAEDAKSNLQREIVEAKGRLEEESIARTNSDRSRKRLEAE 1987


RESULT  12
ECM_HUMAN
ID   ECM_HUMAN      STANDARD;      PRT;  1228 AA.
AC   Q13201;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ENDOTHELIAL CELL MULTIMERIN PRECURSOR.
GN   ECM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 368-376.
RC   TISSUE=Endothelial cells;
RX   MEDLINE=95355440; PubMed=7629143;
RA   Hayward C.P.M., Hassell J.A., Denomme G.A., Rachubinski R.A.,
RA   Brown C., Kelton J.G.;
RT   "The cDNA sequence of human endothelial cell multimerin. A unique
RT   protein with RGDS, coiled-coil, and epidermal growth factor-like
RT   domains and a carboxyl terminus similar to the globular domain of
RT   complement C1q and collagens type VIII and X.";
RL   J. Biol. Chem. 270:18246-18251(1995).
CC   -!- FUNCTION: CARRIER PROTEIN FOR PLATELET (BUT NOT PLASMA) FACTOR
CC       V/VA. MAY PLAY A ROLE IN THE STORAGE AND STABILIZATION OF FACTOR V
CC       IN PLATELETS.
CC   -!- SUBUNIT: MULTIMERIC. COMPOSED OF VARYING SIZED, DISULFIDE-LINKED
CC       MULTIMERS, THE SMALLEST OF WHICH IS A HOMOTRIMER. PROTEOLYSIS OF
CC       THE PROMULTIMERIN IN THE N-TERMINAL REGION, LEADS TO THE MATURE
CC       P155 FORM THAT IS STORED IN PLATELETS.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED BY ENDOTHELIAL CELLS AND
CC       MEGAKARYOCYTES. STORED IN PLATELET ALPHA GRANULES AND ENDOTHELIAL
CC       CELL WEIBEL-PALADE BODIES, FOLLOWING ACTIVATION OF THESE CELLS, IT
CC       IS RELEASED AND ATTACHED TO MEGAKARYOCYTES, PLATELETS, ENDOTHELIUM
CC       AND SUBENDOTHELIUM OF BLOOD VESSELS. NOT FOUND IN PLASMA. FOUND IN
CC       VASCULAR TISSUES SUCH AS PLACENTA, LUNG, AND LIVER.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- PTM: EXTENSIVELY N-GLYCOSYLATED.
CC   -!- DISEASE: DEFICIENCY IN MULTIMERIN DUE TO PROTEOLYTIC DEGRADATION
CC       WITHIN THE PLATELET ALPHA GRANULES IS ASSOCIATED WITH AN AUTOSOMAL
CC       DOMINANT BLEEDING DISORDER (FACTOR V QUEBEC).
CC   -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U27109; AAC52065.1; -.
DR   MIM; 601456; -.
DR   InterPro; IPR000152; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001073; -.
DR   Pfam; PF00008; EGF; 1.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   PROSITE; PS01113; C1Q; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   Signal; Glycoprotein; EGF-like domain; Coiled coil.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1228       ENDOTHELIAL CELL MULTIMERIN.
FT   DOMAIN    1041   1077       EGF-LIKE.
FT   DOMAIN    1119   1149       C1Q.
FT   DOMAIN     333    365       COILED COIL (POTENTIAL).
FT   DOMAIN     400    430       COILED COIL (POTENTIAL).
FT   DOMAIN     503    523       COILED COIL (POTENTIAL).
FT   DOMAIN     580    650       COILED COIL (POTENTIAL).
FT   DOMAIN     675    726       COILED COIL (POTENTIAL).
FT   DOMAIN     819    869       COILED COIL (POTENTIAL).
FT   DOMAIN     189    192       POLY-SER.
FT   DOMAIN     309    313       POLY-GLN.
FT   SITE       186    188       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID  1045   1056       BY SIMILARITY.
FT   DISULFID  1050   1065       BY SIMILARITY.
FT   DISULFID  1067   1076       BY SIMILARITY.
FT   CARBOHYD    21     21       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    97     97       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   114    114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   120    120       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   136    136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   344    344       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   431    431       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   507    507       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   541    541       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   618    618       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   680    680       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   729    729       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   783    783       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   816    816       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   828    828       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   840    840       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   921    921       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   933    933       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   942    942       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   981    981       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1020   1020       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1075   1075       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1228 AA;  138071 MW;  EAA98B0A17E2C4CD CRC64;


  Query Match             6.5%;  Score 133;  DB 1;  Length 1228;
  Best Local Similarity   21.5%;  Pred. No. 5.1;
  Matches   71;  Conservative   68;  Mismatches  111;  Indels   80;  Gaps   17;

Qy      29 YRSLK-KQIEKNIITFNLNLNDILNSRLKKRKYFLDVLESDLMQFKHISSNEYIIEDSFK 87
           |:::|   ::| |   :| :||: |:           ||  : : |   | |:  :   |
Db     330 YQAMKLTLLQKKIDNISLTVNDVRNT--------YSSLEGKVSEDK---SREF--QSLLK 376

Qy      88 LLNSEQKNILLKSYKYIKES---VENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEEK 144
             | |:  |:|::    ::|     :||:   || ::   | ::   :|||| :::|::
Db     377 GLKSKSINVLIRD--IVREQFKIFQNDM---QETVAQLFKTVSSLSEDLESTRQIIQKVN 431

Qy     145 EKFPSSPPTTPPSPAKTDEQKKESKFL------PFLTNIETLYNNLVNKIDDYLINLKAK 198
           |   |                 : ||:      | ||:|  | |::||   :  :  :
Db     432 ESVVSI--------------AAQQKFVLVQENRPTLTDIVELRNHIVNVRQEMTLTCEKP 477

Qy     199 INDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKKDMLGKLLSTG 258
           | :  |:  : |::      |:    :  |:     :|:: | ::    |::  :|||
Db     478 IKELEVK--QTHLE----GALEQEHSRSILY-----YESLNKTLS--KLKEVHEQLLSTE 524

Qy     259 LVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFR--HLDE------------ 304
           |    |   :: :     : :: : |: :|||       |   |: |
Db     525 QVSDQKNAPAAESVSNNVTEYMS-TLHENIKKQSLMMLQMFEDLHIQESKINNLTVSLEM 583

Qy     305 -----REEC-----KCLLNYKQEGDKCEEN 324
                | ||     ||  ::| :    |||
Db     584 EKESLRGECEDMLSKCRNDFKFQLKDTEEN 613
```

```
RESULT 13
MLP1_YEAST
ID   MLP1_YEAST     STANDARD;      PRT;  1875 AA.
AC   Q02455;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   MYOSIN-LIKE PROTEIN MLP1.
GN   MLP1 OR YKR095W OR YKR415.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=93247549; PubMed=8483450;
RA   Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT   "A new yeast gene with a myosin-like heptad repeat structure.";
RL   Mol. Gen. Genet. 237:359-369(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94205265; PubMed=8154186;
RA   Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA   Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT   "The complete sequence of a 15,820 bp segment of Saccharomyces
RT   cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RT   new open reading frames.";
RL   Yeast 9:1349-1354(1993).
CC   -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC       REPAIR.
CC   -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC   -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L01992; AAA34783.1; -.
DR   EMBL; X73541; CAA51948.1; -.
DR   EMBL; Z28320; CAA82174.1; -.
DR   PIR; S38173; S38173.
DR   SGD; S0001803; MLP1.
KW   Myosin; Heptad repeat pattern; Coiled coil; DNA repair.
FT   DOMAIN       69    487       COILED COIL (POTENTIAL).
FT   DOMAIN      531   1678       COILED COIL (POTENTIAL).
FT   DOMAIN     1834   1866       COILED COIL (POTENTIAL).
FT   CONFLICT    301    301       R -> A (IN REF. 1).
SQ   SEQUENCE   1875 AA;  218455 MW;  683A0D34C9066867 CRC64;


  Query Match             6.5%;  Score 132.5;  DB 1;  Length 1875;
  Best Local Similarity   21.6%;  Pred. No. 8.5;
  Matches   86;  Conservative   83;  Mismatches  126;  Indels  103;  Gaps   23;

Qy       7 DNILSGFENEYDVIYLKPLAGVYRSLKKQIEKNIITFNLNLN---DILNSRLKKRKYFLD 63
           ||     ::  | |:||      |  |: ::  | || : |     |:||  |:|    |
Db     374 DNENLSAKSSSDFIFLKKQLIKERRTKEHLQNQIETFIVELEHKVPIINS-FKER---TD 429

Qy      64 VLESDL----MQFKHISS-----------NEYII--EDSFKLLNSEQKNILLK-SYKYI 104
             :||::|    :  :| |:            |: ::  |:  : |  :: ::  :  |  |
Db     430 MLENELNNAALLLEHTSNEKNAKVKELNAKNQKLVECENDLQTLTKQRLDLCRQIQYLLI 489

Qy     105 KESVENDIK--FAQEGISYYEKVLAKYKDDLESI-----KKVIKEE----------KEKF 147
             || || |    :| | : : ::   ::|  :|     :||: |           :||
Db     490 TNSVSNDSKGPLRKEEIQFIQNIM---QEDDSTITESDSQKVVTERLVEFKNIIQLQEKN 546

Qy     148 PSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLVNKIDDYLINLK-------AKIN 200
                    | : ::|:||      ::: : :  ||:  : :| ||       ::|
Db     547 AELLKVVRNLADKLESKEKKSK-----QSLQKIESETVNEAKEAIITLKSEKMDLESRIE 601

Qy     201 DCNVEKDE-----------AHVKITKLSDLKA-IDDKI-DL--------FKNTNDFEAI 238
           :   | :|            ::| | :|:: |  :: :: ||         ::| :   :
Db     602 ELQKELEELKTSVPNEDASYSNVTIKQLTETKRDLESQVQDLQTRISQITRESTENMSLL 661

Qy     239 KKLIND--DTKKDM---LGK-----LLSTGLVQIFPNTI-ISKLIEGKFQDMLNISQHQC 287
           | | |  |:| |:   |||     :|:    ::  ||: ::|    : :   :  |:
Db     662 NKEIQDLYDSKSDISIKLGKEKSSRILAEERFKLLSNTLDLTKAENDQLRKRFDYLQNTI 721

Qy     288 VKKQCPENSGCFRHLDEREECKC--------LLNYKQE 317
           :|    ::|     |:|   ||          ||| |:|
Db     722 LK----QDSKTHETLNEYVSCKSKLSIVETELLNLKEE 755


RESULT 14
NUF1_YEAST
ID   NUF1_YEAST     STANDARD;      PRT;   944 AA.
AC   P32380;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   NUF1 PROTEIN (SPINDLE POLY BODY SPACER PROTEIN SPC110).
GN   NUF1 OR SPC110 OR YDR356W OR D9476.3.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=92176232; PubMed=1541631;
RA   Mirzayan C., Copeland C.S., Snyder M.;
RT   "The NUF1 gene encodes an essential coiled-coil related protein that
RT   is a potential component of the yeast nucleoskeleton.";
RL   J. Cell Biol. 116:1319-1332(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94064779; PubMed=7503995;
RA   Kilmartin J.V., Dyos S.L., Kershaw D., Finch J.T.;
RT   "A spacer protein in the Saccharomyces cerevisiae spindle poly body
RT   whose transcript is cell cycle-regulated.";
RL   J. Cell Biol. 123:1175-1184(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA   Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA   Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA   Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA   Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA   Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,
RA   Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEOSKELETON. POTENTIAL
CC       ROLE IN CROSSLINKING FILAMENTS OR ANCHORING OTHER MOLECULES. IT
CC       IS ESSENTIAL FOR GROWTH.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; TIGHTLY ASSOCIATED WITH THE
CC       NUCLEUS. IT IS PRESENT IN A GRANULAR PATTERN THAT EXCLUDES THE
CC       NUCLEOLUS.
CC   -!- PTM: MAY BE REGULATED BY PHOSPHORYLATION EVENTS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z11582; CAA77668.1; -.
DR   EMBL; X73297; CAA51733.1; -.
```

```
DR   EMBL; U28372; AAB64791.1; -.
DR   PIR; S26710; S26710.
DR   PIR; S34288; S34288.
DR   SGD; S0002764; NUF1.
KW   Coiled coil; Nuclear protein; Phosphorylation; Heptad repeat pattern.
FT   DOMAIN      164    791       COILED COIL.
FT   DOMAIN       54     59       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      726    731       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      742    747       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      731    944       ARG/TYR-RICH.
SQ   SEQUENCE   944 AA;  111781 MW;  04FAA074BB8A0BC8 CRC64;


  Query Match             6.4%;  Score 132;  DB 1;  Length 944;
  Best Local Similarity   23.7%;  Pred. No. 4.4;
  Matches   85;  Conservative   64;  Mismatches  142;  Indels   68;  Gaps   18;

Qy     27 GVYRSLKKQIEKNIITFNLNLNDI-LNSRLKKRKYFLDVLESD--LMQFKHISSNEYIIE 83
          |   ||  |: :  :   ||  |  | || :: |   | |: |  : : : :| :| ||:
Db    381 GKLASLMAQLTQ--LESKLNQRDSQLGSREEELKKTNDKLQKDIRIAREETVSKDERIID 438

Qy     84 ----------DSF--KLLNSEQKNIL---LKSYKYIKESVENDIKFAQEGISYYEKVLAK 128
                    | |  |  :|| | |    |:|   : : :|||:| |||  |  || | :
Db    439 LQKKVKQLENDLFVIKKTHSESKTITDNELESKDKLIKILENDLKVAQEKYSKMEKELKE 498

Qy    129 YKDDLESIKKVIKEEK----EKFPSSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNL 184
           : : :  :  :::||    ||  :          | ::    :  :    | |    :|
Db    499 REFNYKISESKLEDEKTTLNEKISNLAAENSQLKNKIEDNSTATHHMK--ENYEKQLESL 556

Qy    185 VNKIDDYLINLK---AKINDCNVEKDEAHVKIT--KLSDLKAIDDKI-DLFKN----TND 234
            |::|  : |    || :  :   |  |::  :  |:|  |::| || :|    ::
Db    557 RKDIEEYKESAKDSEDKIEELKIRIAENSAKVSEKRSKDIKQKDEQISDLTQNLKLQEDE 616

Qy    235 FEAIKKLINDDTKKDMLGKLLSTGLVQIFPNTIISKLIEGKFQDMLNISQHQCVKKQCPE 294
            ::| :| |  |||              | : |:  :   |  ||:       |
Db    617 ISSLKSII-DRYKKDF-------------NQLKSE--QSNIQHDLNLQILNLENKL--- 656

Qy    295 NSGCFRHLDEREECKCLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKK 353
                ::  :| | |    ::  | |   |     ||   : |   ||::| | :::
Db    657 -------IESEDELKSL----RDSQKIEIE-NWKRKYNNLSLENDRLLTEKESASDKER 703


RESULT  15
CIN8_YEAST
ID   CIN8_YEAST     STANDARD;      PRT;  1038 AA.
AC   P27895;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   KINESIN-LIKE PROTEIN CIN8.
GN   CIN8 OR KSL2 OR YEL061C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=92317149; PubMed=1618897;
RA   Hoyt M.A., He L., Loo K.K., Saunders W.S.;
RT   "Two Saccharomyces cerevisiae kinesin-related gene products required
RT   for mitotic spindle assembly.";
RL   J. Cell Biol. 118:109-120(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA   Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA   Chung E., Duncan M., Guzman E., Hartzell G., Hunicke-Smith S.,
RA   Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA   Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA   Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA   Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1021-1038 FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=95172238; PubMed=7867803;
RA   Rousselet G., Simon M., Ripoche P., Buhler J.M.;
RT   "A second nitrogen permease regulator in Saccharomyces cerevisiae.";
RL   FEBS Lett. 359:215-219(1995).
RN   [4]
RP   CHARACTERIZATION.
RC   STRAIN=S288C;
RX   MEDLINE=92354062; PubMed=1643659;
RA   Saunders W.S., Hoyt M.A.;
RT   "Kinesin-related proteins required for structural integrity of the
RT   mitotic spindle.";
RL   Cell 70:451-458(1992).
CC   -!- FUNCTION: REQUIRED FOR ASSEMBLY OF THE MITOTIC SPINDLE. INTERACT
CC       WITH SPINDLE MICROTUBULES TO PRODUCE AN OUTWARDLY DIRECTED
CC       FORCE ACTING UPON THE POLES. FOLLOWING SPINDLE ASSEMBLY, CIN8 AND
CC       KIP1 APPARENTLY ACT TO OPPOSE A FORCE THAT DRAWS SEPARATED POLES
CC       BACK TOGETHER. THIS FORCE SEEMS TO BE MEDIATE BY KAR3.
CC   -!- SUBCELLULAR LOCATION: SPINDLE MICROTUBULES THAT LIE BETWEEN THE
CC       POLES.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z11859; CAA77885.1; -.
DR   EMBL; M90522; AAA34496.1; -.
DR   EMBL; U18795; AAB65026.1; -.
DR   EMBL; X79105; CAA55722.1; -.
DR   PIR; B42641; B42641.
DR   HSSP; P17119; 3KAR.
DR   SGD; S0000787; CIN8.
DR   InterPro; IPR001752; -.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
KW   Mitosis; Cell cycle.
FT   DOMAIN       72    553       MECHANOCHEMICAL (MOTOR) (BY SIMILARITY).
FT   DOMAIN      554    657       COILED COIL (POTENTIAL).
FT   DOMAIN      904    942       COILED COIL (POTENTIAL).
FT   NP_BIND     166    173       ATP (BY SIMILARITY).
FT   CONFLICT    254    254       D -> A (IN REF. 1).
FT   CONFLICT    831    831       Q -> H (IN REF. 1).
SQ   SEQUENCE   1038 AA;  117999 MW;  3A1FD7003EF89FBC CRC64;


  Query Match             6.4%;  Score 130.5;  DB 1;  Length 1038;
  Best Local Similarity   25.3%;  Pred. No. 5.7;
  Matches   79;  Conservative   54;  Mismatches  126;  Indels   53;  Gaps   15;

Qy     29 YRSLKKQIEKNIITFNLNLNDILNSRLKKRKYFLDVLESDLMQFKHISSNE--YIIEDSF 86
          | |  | |:      :  : |||   :|     |  ::|||:  |   | |  |: :| :
Db    508 YASKAKNIKNKPQLGSFIMKDIL---VKNITMELAKIKSDLLSTK---SKEGIYMSQDHY 561

Qy     87 KLLNSEQKNILLKSYKYIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEEKEK 146
          | |||:     |:|||   :  : :|:      |   |     || |:| |: |: :  :
Db    562 KNLNSD-----LESYKNEVQECKREIE------SLTSKNALLVKDKLKS-KETIQSQNCQ 609

Qy    147 FPSSPPTTPPSPAKTDEQKKESKFLPFLTNIE-TLYNNLVNKIDDY----LINLKAKIND 201
            |   |     |: |:| |        | || : :|| : |: :     | : | :  |
Db    610 IESLKTTIDHLRAQLDKQHK--------TEIEISDFNNKLQKLTEVMQMALHDYKKRELD 661
```

```
Qy     202 CNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDDTKKDMLGKLLSTGLVQ 261
           | :| | |  |||  ::| :   : |  ||   |:| :  |     ||:
Db     662 LN-QKFEMH--ITK--EIKKLKSTLFLQLNTMQQESILQETNIQPNLDMI---------- 706

Qy     262 IFPNTIISKLIEGKFQDMLNISQHQCVKKQCPENSGCFRHLDEREECKCLLNYKQEGDKC 321
             | :::  :    |:   :    ||||   |:   |  : |: :    :::::
Db     707 --KNEVLT--LMRTMQEKAELMYKDCVKKILNESPKFFNVVIEKIDI-IRVDFQKFYKNI 761

Qy     322 EENPNPTCNENN 333
           || :    |||
Db     762 AENLSDISEENN 773


Search completed: August  8, 2001, 12:32:34
Job time: 202 sec
```